

Bailor, Edwin M

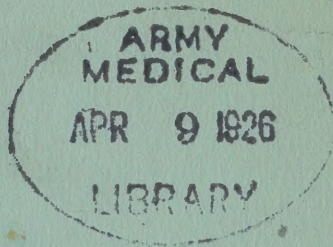
CONTENT AND FORM IN TESTS OF INTELLIGENCE

INDEXED

MK

BY

EDWIN MAURICE BAILOR



SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIRE-
MENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY IN
THE FACULTY OF PHILOSOPHY, COLUMBIA UNIVERSITY

Amer

Published by

Teachers College, Columbia University

New York City

1924

CONTENT AND FORM IN TESTS OF INTELLIGENCE

BY

EDWIN MAURICE BAILOR

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIRE-
MENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY IN
THE FACULTY OF PHILOSOPHY, COLUMBIA UNIVERSITY

Published by
Teachers College, Columbia University
New York City
1924

COPYRIGHT, 1924, TEACHERS COLLEGE, COLUMBIA UNIVERSITY

J. J. LITTLE & IVES COMPANY, NEW YORK, 10-17-24-100

FOREWORD

Development of mental and educational tests is only one expression of a newer movement in education which, in my opinion, promises more beneficial results than any of the successive waves of educational enthusiasm of the past century. The significance of group thinking on educational problems, of shared responsibility for their solutions, and of increasing co-operation, especially in research, which has developed within the past decade does not appear to be sufficiently recognized. The complex problems in education have been too difficult for any individual or single group of individuals to solve. If solutions of these problems are to be found, and they will be, increasingly professionalized attitudes, combined with coöperation, should result in valuable contributions. Research in educational psychology and all of its fields is dependent upon coöperation; especially is this true of work in tests and measurement. It is particularly true in the present study, for progress in it as well as its inception has been made possible only through the coöperation of many.

It is with considerable regret that I have come to realize that it is impracticable to treat these data with the more finished statistical analysis which is being accorded to studies in physics, biology, anthropology, and other sciences. This practice is becoming increasingly more necessary in educational and psychological research, and without it a study lacks desired completeness. First, for example, it would be desirable that the types of distribution found here be studied, for they might disclose discriminating values of the tests at different levels of difficulty. Such a study, also, would have determined the justification of the use, as frequently is done, of certain formulae without testing to determine whether actual conditions fulfil the assumptions made in their derivations. Second, the probable errors of coefficients of partial correlation and of coefficients corrected for attenuation have not been computed. In the latter case their computation is impossible without dupli-

cation of work. Most of the correlation tables, except those for reliability, have been plotted by the differences of deviations method which the use of the Toops formulae permits. This method has made it impossible to compute the higher mixed moments which are necessary for determining probable errors of corrected coefficients. Finally, it would be desirable, for completeness, that all intercorrelations of the thirty tests be computed, but the labor required for such computations does not seem to be justified, at least in the present study, by the apparently little additional value which might result.

It has seemed best, then, to forego some of the steps in the study which would have improved it, for the practical implications of the results are believed to be sufficiently well established. The satisfaction of practical demands for early publication seems more reasonable than indefinite delay out of respect for theoretical perfection.

It is impossible to acknowledge in detail the obligations for which the writer is indebted. I wish, first, to acknowledge my obligation to Professor E. L. Thorndike for his help in the selection of this problem and for his constant guidance and inspiring influence. For the original records of the one thousand thirty-nine cases of the miscellaneous group and the four hundred eighty-nine cases of the age-control group, I am indebted to the Commonwealth Fund and the Carnegie Corporation, whose grants to the Institute of Educational Research of Teachers College made the testing and scoring of these subjects possible. For the use of these data I thank the Institute of Educational Research. To the personnel of the Institute I owe my thanks for wholehearted coöperation and repeated daily courtesies.

To Professor Henry A. Ruger I owe special thanks. During the one and one-half years while I have been assisting him in his classes he has without exception been helpful, considerate, and encouraging. He has given freely of his time, and his judgment and guidance have been helpful, sound, and constant. To him I am indebted for my general statistical training as well as for advice on special problems as they have been confronted in this study. But for the greater inspiration which he, as my teacher, has had upon me, his pupil, I am most greatly indebted.

I wish to acknowledge also the help of Professor Godfrey H. Thomson; of members of the psychological seminar of Teachers

College and of the graduate students who assisted in giving tests; the coöperation of Mr. Nelson C. Smith, supervising principal of the Leonia, New Jersey, Public Schools, of his corps of teachers and of the pupils of his school; of Mr. Robert Burns, principal of the High School, Grantwood, New Jersey; and the help of those unknown thousands in "City I" as a result of whose coöperation the data, studied here, were obtained.

But most of all I am indebted to my wife, Jane Galt Bailor, who has shared the joys and work with me from the beginning of this study to its publication.

E. M. B.

April, 1924.

CONTENTS

CHAPTER	PAGE
I. THE PROBLEM	I
II. THE TESTS	4
General Description of the Tests	4
Validity and Reliability of the Tests	8
Weighting of the Test Scores	9
Correlation Methods Employed	10
III. THE SUBJECTS	12
The Major Study	12
Selection of Subject	12
Selection of Records	14
Age-Control Group	14
The Time-Controlled Test	15
Desirability of Control Tests	15
Preparation of Papers	15
Selection of Subjects and Administration of Tests	16
Summary of Material and Method	17
IV. EFFECT OF DIFFERENCES IN CONTENT	18
Method of Grouping, and Results	18
Significance of Correlation Coefficients	25
The Predictive Index	26
Correspondence of Results of Groups of Tests Having Differences in Content	29
Test Groups	29
Differences Between Single Tests	32
Tests of the Analogies Form	32
Tests of the Generalization Form	33
Tests of the Completion Form	33
Tests Different in Both Content and Form	33
V. EFFECT OF DIFFERENCES IN FORM	35
Method of Grouping and Results	35
Correspondence of Results of Test-Groups Hav- ing Differences in Form	36
Numerical Tests Having Differences in Form	41

CHAPTER	PAGE
Verbal Tests Having Differences in Form . . .	41
Spatial Tests Having Differences in Form . . .	43
Correlation of Test Groups with a Composite . . .	43
VI. THE CONTROL GROUP	46
Effect of Fluid Time Allowance	46
Effect of Age	50
Constancy of Correlation Coefficients	50
Comparison with Terman Group Test Results	52
VII. CONCLUSIONS AND SUMMARY	55
Summary	62
APPENDIX	
I. Transmutation Tables I and II and Representative Correlation Table	65
II. Modified Instructions for Time-Controlled Test	68
III. Bibliography	73

TABLES

NUMBER	PAGE
1. STANDARD DEVIATIONS AND ARITHMETIC MEANS OF FIFTEEN COMPONENT TESTS	6
2. RELIABILITY COEFFICIENTS OF THE FIFTEEN COMPONENT TESTS	9
3. POPULATION OF "CITY I".	13
4. COMPARATIVE STANDARD DEVIATIONS AND MEANS OF SINGLE TESTS GROUPED ACCORDING TO CONTENT	18
5. STANDARD DEVIATIONS AND ARITHMETIC MEANS OF TEST-GROUPS DIFFERING IN CONTENT (MISCELLANEOUS GROUP AND AGE-CONTROL GROUP)	19
6. COEFFICIENTS OF INTERCORRELATION BETWEEN TESTS GROUPED ON THE BASIS OF CONTENT	20
7. INTERCORRELATIONS OF ANALOGIES TESTS — SPATIAL, GRAMMATICAL, AND VERBAL IN CONTENT BUT SIMILAR IN FORM	22
8. INTERCORRELATIONS OF GENERALIZATION TESTS—VERBAL AND NUMERICAL IN CONTENT BUT SIMILAR IN FORM	23
9. INTERCORRELATIONS OF COMPLETION TESTS—VERBAL AND NUMERICAL IN CONTENT BUT SIMILAR IN FORM	24
10. CORRELATIONS BETWEEN TESTS DIFFERENT IN FORM AND DIFFERENT IN CONTENT	24
11. PREDICTIVE INDICES FOR CERTAIN CORRELATION COEFFICIENTS	28
12. COMPARATIVE STANDARD DEVIATIONS AND MEANS OF SINGLE TESTS GROUPED ACCORDING TO FORM	35
13. STANDARD DEVIATIONS AND MEANS OF TEST-GROUPS DIFFERING IN FORM (MISCELLANEOUS GROUP AND AGE-CONTROL GROUP)	36
14. COEFFICIENTS OF INTERCORRELATION BETWEEN TESTS GROUPED ON THE BASIS OF FORM	37
15. INTERCORRELATIONS OF NUMERICAL TESTS—TESTS OF DIFFERENT FORM BUT ALL OF NUMERICAL CONTENT	41
16. CORRELATIONS BETWEEN TESTS OF DIFFERENT FORM BUT BOTH OF VERBAL CONTENT	42
17. CORRELATIONS BETWEEN TESTS OF DIFFERENT FORM BUT BOTH OF SPATIAL CONTENT	42
18. COMPARATIVE STANDARD DEVIATIONS AND MEANS OF GROUPS OF TESTS DIFFERING IN CONTENT (TIME-CONTROL GROUP)	47
19. COMPARATIVE STANDARD DEVIATIONS AND MEANS OF GROUPS OF TESTS DIFFERING IN FORM (TIME-CONTROL GROUP)	47

NUMBER	PAGE
20. INTERCORRELATIONS OF TIME-CONTROLLED TEST AND COMPARISON WITH MISCELLANEOUS GROUP	48
21. MEAN DEVIATIONS AND MEAN PROBABLE ERRORS OF CROSS-CORRE- LATION COEFFICIENTS	51
22. CORRELATIONS BETWEEN SCORES ON TERMAN GROUP TEST AND SCORES ON I. E. R. TEST	53
APPENDIX	
23. TRANSMUTATION TABLE I: GROUPING TABLE FOR DETERMINING TRANS- MUTED SCORES FROM RAW SCORES ON GROUPS OF TESTS	65
24. TRANSMUTATION TABLE II: GROUPING TABLE FOR DETERMINING TRANSMUTED SCORES FROM RAW SCORES ON COMPONENT TESTS .	66
25. REPRESENTATIVE CORRELATION TABLE—WORDS SERIES A; AND NUM- BERS SERIES B	67

CONTENT AND FORM IN TESTS OF INTELLIGENCE

CHAPTER I

THE PROBLEM

The purpose of this study is to determine the correspondence between mental abilities as measured by the results of certain tests of the same subjects when tests having different types of content and differences in form are used. If differences in content and form of so-called tests of intelligence have no material effect upon relative results, these properties of tests are not important in test construction and in their selection. If differences either in content or in form, or both, materially affect test results, the fact is of importance in leading toward improvement in intelligence testing. The problem should be of importance first to the maker of tests, for it is he who must select and weight the materials used in each test and define or evaluate their results. The user of tests also is concerned, because if the particular content and form of a test do have marked effect upon the results obtained, selection of tests must be made on the basis of which ones are best fitted, by reason of their content and construction, to measure most accurately those abilities which it is desired to measure. The methods and materials used in the study of this problem are described in the following chapters.

A statement of the meaning of the terms used in this study is appropriate at this point. The word *test* is used to denote a group of test elements, similar in content and form, such as the arithmetical problems test, the sentence completion test, and so on.

The use of the terms *content* and *form* departs slightly from the following definition given by Hart and Spearman (1912, page 52): “. . . by form is meant the kind of mental operation as discrimination, observation, inference, etc., while content denotes the different sorts of data, as color, shape, number, etc.,

submitted to such operations." As used in this study *content* will denote the kinds of data which are defined, but will be limited generally to three types: (1) *verbal*, (2) *numerical*, and (3) *spatial*. The verbal type is concerned with word meanings; an equivalent term for it is *words*. The numerical type is concerned with number relations and is spoken of simply as *numbers*; the spatial type involves line pictures and geometrical relations and will be called *space*.

The use of the word *form* departs more widely from the definition quoted. Hart and Spearman indicate that the word is used to designate some particular mental function of presumably more or less integral nature which the test is supposed to measure. Inasmuch as so little is known of the nature of the intellect and since there is so much doubt about the integral nature of mental functions, grave danger seems to lurk in any such basis of classification.

An objective basis of classification seems more sound. The word *form* is therefore used in the present study to denote the objective mold in which the test is cast; for example, whether it requires the addition of elements to complete a series or a sentence, as in a completion test; whether it involves the selection of a word, picture, or number which has the greatest similarity to a given group of elements, as in a generalization or similarities test; or, whether it is a proportion in which a fourth term is to be selected which has the same relation to a third as a second has to a first, as in an analogies test. More simply by *form* is meant the objective arrangement in which the problem is presented. Four such classes will be considered here and will be called *completion tests*, *analogies tests*, *generalization tests*, and a *composite* which is made up of five tests, all of which are of still different objective composition. By differences in *form*, then, are meant such differences as exist between completion tests, analogies tests, and generalization tests.

In the use of these terms no assumption is made regarding the nature of the mental processes involved in the respective types of tests presented. There is no assumption that the abilities tested by words, for example, are either "multifocal, intermediate or unifocal" (Hart and Spearman, 1912, pages 52, 53); nor is any assumption made regarding the nature of abilities which are tested by the numerical or spatial tests. Similarly,

there is no assumption on the part of the writer regarding the abilities measured by the completion, the analogies, or the generalization tests, or by the composite test; no "faculties" as such are presupposed in this characterization of the tests. In every case the terms words, numbers, and space, and generalization, completion, analogies and composite refer to the specific tests or groups of tests, or their results, as they have been described and determined in this study, except where the context clearly shows a recognized and more general use.

The attempt here has been made to make the classification objective. It may therefore lack psychological justification for the distinctions drawn. But who knows what are the abilities involved in each type? In a practical situation if intelligence tests are to be used they must be characterized by having some type of *content* and some kind of *form* in the sense used here. It is the purpose of this problem to study the relations between results of certain tests which differ in content and in form. Whatever the conclusions may be which are reached in this study, they can be generalized only to the extent to which the definitions and meanings used here are applicable.

CHAPTER II

THE TESTS

GENERAL DESCRIPTION OF THE TESTS

The data for the major part of this study are those which were obtained by Thorndike in his study of "Mental Discipline in High School Studies" ('24 A and '24 B)¹. The original description (Thorndike, '22) should be consulted for a detailed explanation of the nature of the tests and the principles underlying their selection. To summarize briefly, tests were selected which might be of greatest value as measures of "ability to think with symbols, to understand and apply generalizations, to discern and use relations and to select essential facts or elements, and to organize facts for a purpose, in James' phrase to 'think things together.'" Because they were so selected they were called "Tests of Selective and Relational Thinking and Generalization and Organization." The materials are prepared in duplicate batteries or series, Series A and Series B, each consisting of two booklets, eight pages each, bearing these respective titles. In addition a *Practice Booklet* of eight pages was prepared. This booklet was presented first to the subjects as an introduction to the nature and method of the real examination, but its results were not scored. The tests of the two booklets, Selective and Relational Thinking and Generalization and Organization, compose the real examination. While these tests were not originally selected and arranged for the study named (Thorndike, '22), they were offered as a useful measure of improvement in the mental abilities tested.

A further discussion of the value of the tests lies outside of the purpose of this study. However, it is worth notice, in passing, that an interesting implication regarding one characteristic of so-called intelligence immediately arises when one considers that the fifteen tests (Test 3 excepted) of which the battery is composed, are tests which have been framed and

¹ The references cited in this study are given in detail in the Bibliography, pp. 73, 74.

used more or less generally by recognized psychologists in their so-called intelligence tests. It is natural, then, that there should be marked correspondence between the results of this battery and a well-recognized standardized test. Thorndike ('24 A, p. 4) cites that for two hundred four cases, the correlation between the various forms of the battery under consideration and the Terman Group Tests, Forms A and B, is .82. As will be seen later, results of this study disclose that the reliability coefficient for Series A and B of this examination given at an interval of one year is .8198 for the miscellaneous group of one thousand thirty-nine cases studied and .8633 for the group of four hundred eighty-nine boys sixteen years or more but less than seventeen years of age. Since, therefore, its correlation with a recognized test is nearly as high as its own reliability, it can to that extent properly be called a test of intelligence. The fact is to be expected when it is noticed that the tests are the familiar ones found in Army Alpha, the Terman Group, the Army Beta, and others, and in adaptations of such special tests as the Woodworth-Wells absurdities, the Thorndike-Wylie opposites, Rogers number-series completion, the Briggs grammatical analogies, and the Otis generalization or similarities tests involving common features in words, numbers, and pictures. It should give a measure, therefore, of so-called general intelligence or whatever is measured by the Stanford-Binet, Otis, National, Terman, Haggerty, Thorndike, and similar tests.

Examination of the tests themselves, together with instructions for giving them, would give a more exact idea of their nature.¹ To secure clearness when referring to the fifteen respective tests in the series, numbers will be used. Since the series of Selective and Relational Thinking were given first, the eight tests comprising this part of the examination will be given their original serial numbers from one to eight, inclusive, and the seven tests of the Generalization and Organization booklet will be given the numbers from nine to fifteen, inclusive, in the serial order in which they occur. For example, Test 4 of the Selective and Relational Thinking booklet, the Thorndike-Wylie

¹ A splendid description of these test materials appears in the *Journal of Educational Research*, Volume V, No. 4 (April, 1922), pages 269-279, under the title, "Instruments for Measuring the Disciplinary Values of Studies," by E. L. Thorndike; hence, it is not repeated here. Copies of the tests and directions for their administration are obtainable from the Institute of Educational Research, Teachers College, Columbia University, New York City.

TABLE 1 A

STANDARD DEVIATIONS AND ARITHMETIC MEANS OF FIFTEEN COMPONENT TESTS
Miscellaneous Group (1039)

Number of Minutes Allot- ted	Test Num- ber	Name of Test	Series A (1922)		Series B (1923)	
			S. D.	Mean	S.D.	Mean
10	1	Arithmetical Problems	8.2304	39.2048	8.1256	37.1300
4	2	Absurdities (Thorndike, after Woodworth)	3.6608	8.8402	3.3296	9.3985
4	3	Line Arrangement (Thorndike)	3.2020	3.2613	4.5362	7.6992
5	4	Opposites (Wylie)	6.5466	24.5096	7.1586	27.3941
4	5	Number Series Completion (Thorndike, after Rogers)	4.5950	8.2993	4.6216	10.1222
4	6	Spatial (Geometrical) Analogies (Thorndike)	4.3988	5.6737	4.7030	12.3956
4	7	Grammatical Analogies (Thorndike, Selection and Extension from Briggs)	4.4038	6.1261	4.6358	7.4754
4	8	Verbal Analogies (Thorndike, after Army Alpha, after Woodworth-Wells)	4.9472	10.5958	4.9904	12.5515
2	9	Moral Judgment (Pressey)	3.9410	12.1915	4.3886	12.3455
3	10	Verbal Generalization (Thorndike, after Otis)	5.2996	8.7459	5.4914	8.5438
3	11	Geometrical Generalization (Thorndike, after Otis)	3.2314	6.1102	3.3314	6.1882
3	12	Numerical Generalization (Thorndike, after Otis)	3.2196	4.9418	4.5178	8.4171
10	13	Trabue Completion J and L (or K and M)	8.8737	28.4393	8.8314	31.5895
3	14	Cutting Up Surfaces (Thorndike, after Army Beta)	2.4667	8.5173	2.2726	9.4701
12	15	Disarranged Arithmetical Equations (Thorndike)	6.9285	14.1266	8.3913	16.7137
		TOTAL		189.5874		217.4344

opposites test, will be designated by its original number and so will be called Test 4; Test 5 of the Generalization and Organization booklet, the Trabue completion test, will be designated as Test 13. A brief reference to the arrangement of these tests should make plain the use of numbers for reference to the component tests of the series.

Although this battery of tests was not originally developed for the present study, it appears to be particularly suitable for it because Tests 1, 6, 7, and 8 are all of the analogies or mixed relations form, but with spatial, grammatical, and word-meaning content. Similarly, Tests 10, 11, and 12 are all of the

TABLE 1 B

STANDARD DEVIATIONS AND ARITHMETIC MEANS OF FIFTEEN COMPONENT TESTS
Age-Control Group (489)

Number of Minutes Allot- ted	Test Num- ber	Name of Test	Series A (1922)		Series B (1923)	
			S.D.	Mean	S.D.	Mean
10	1	Arithmetical Problems.....	8.630	40.0612	7.523	36.0544
4	2	Absurdities (Thorndike, after Wood- worth).....	3.732	8.9198	3.321	9.5070
4	3	Line Arrangement (Thorndike)....	3.361	3.7249	4.654	7.9274
5	4	Opposites (Wylie).....	7.184	25.1379	7.502	28.2729
4	5	Number Series Completion (Thorn- dike, after Rogers).....	4.708	7.9610	4.599	9.8466
4	6	Spatial (Geometrical) Analogies (Thorndike).....	4.376	5.3640	4.822	12.5828
4	7	Grammatical Analogies (Thorn- dike, Selection and Extension from Briggs).....	4.246	6.4182	4.512	7.4458
4	8	Verbal Analogies (Thorndike, after Army Alpha, after Woodworth- Wells).....	5.009	10.7626	4.968	12.3210
2	9	Moral Judgment (Pressey).....	3.938	12.4274	4.154	13.1881
3	10	Verbal Generalization (Thorndike, after Otis).....	5.554	8.7340	5.597	8.6564
3	11	Geometrical Generalization (Thorn- dike, after Otis).....	3.104	5.7883	3.168	5.9458
3	12	Numerical Generalization (Thorn- dike, after Otis).....	3.147	4.9683	4.585	8.3241
10	13	Trabue Completion J and L (or K and M).....	9.438	29.6532	9.293	32.0397
3	14	Cutting Up Surfaces (Thorndike, after Army Beta).....	2.257	8.6635	2.177	9.6431
12	15	Disarranged Arithmetical Equa- tions (Thorndike).....	7.083	14.3343	8.587	17.2299
		TOTAL.....		102.9185		218.9850

generalization or common element form used by Otis, but differ in content having, respectively, word-meaning, spatial, and number content. Thus they give promise of offering evidence as to whether differences in content and form, as properties of tests, are significant in their respective effects upon test results. Test 5 is a number-completion and Test 13 is a Trabue sentence completion test. Tests 2, 3, 4, 7, 8, 9, 10, and 13 have been grouped as being primarily word-meaning tests; the numerical tests are Tests 1, 5, 12, and 15; and Tests 3, 6, 11, and 14 form a group, the content of which is dominantly spatial. It is for-

tunate that the tests of different content were so satisfactorily mingled; there appears no valid reason to suppose that any one group had any marked advantage due to position in the series.

Series B of the tests appears in all essential respects similar to Series A; it is considered a duplicate. Results from its use, as described in the reference cited (Thorndike, '24 A, page 11), indicate that it is approximately two and one-half points easier than Series A. This minor difference should have a negligible effect upon the results of correlation studies of the tests when studied both singly and in the groups later described.

The scoring, as used in the original study, is accepted in the present one. The original plan of scoring was devised (Thorndike, '24 A, page 6) from a preliminary sampling of booklets so as to require a minimum of time and still give reasonable weights to each of the fifteen tests and especially to the total scores in the word-meaning (Words), numerical (Numbers), and spatial (Space) composites. The comparative standard deviations and means are given in Table 1 A and 1 B.

VALIDITY AND RELIABILITY OF THE TESTS

In regard to the validity of the battery of tests as a measure of intelligence only passing comment is necessary. It has already been stated (page 4) that of the fifteen tests all except one have been more or less widely used by recognized psychologists as tests of intelligence. Thorndike ('24, page 4) states that the results of Series A (1922) of this test correlate with the Terman Group Test, Forms A and B being used, $.82 \pm .02$ for two hundred nine (209) cases.

The reliability of the battery, as indicated by the coefficients of self-correlation of the total of the fifteen tests of Series A (1922) and Series B (1923), as determined in this study, is $.8198 \pm .0068$ for the miscellaneous group of one thousand thirty-nine (1039) cases, and $.8633 \pm .0077$ for the age-control group of four hundred eighty-nine (489) cases. The standard deviations and means of these are:

		S.D.	Mean
Miscellaneous Group (1039 Cases)	Series A (1922).....	41.8480	192.0048
	Series B (1923).....	46.0528	223.8976
Age-Control Group (489 cases)	Series A (1922).....	43.5040	198.0400
	Series B (1923).....	47.7760	226.0160

TABLE 2

RELIABILITY COEFFICIENTS OF THE FIFTEEN COMPONENT TESTS

Self-Correlation of Respective Tests of Series A (1922) with Their Duplicates of Series B (1923)

Test Number	Miscellaneous Group (1039 Cases)		Age-Control Group 16 Years (489 Cases)	
	<i>r</i>	P.E.	<i>r</i>	P.E.
1.....	.5769	.0140	.6560	.0166
2.....	.2035	.0201	.2535	.0285
3.....	.4762	.0162	.3380	.0270
4.....	.5265	.0151	.6010	.0195
5.....	.4885	.0159	.5089	.0226
6.....	.4840	.0160	.3049	.0277
7.....	.6562	.0138	.6799	.0164
8.....	.5765	.0140	.6217	.0187
9.....	.2294	.0198	.1163	.0301
10.....	.5068	.0155	.4976	.0229
11.....	.3664	.0181	.4413	.0246
12.....	.2458	.0197	.2906	.0280
13.....	.5009	.0157	.5637	.0208
14.....	.3615	.0182	.4010	.0256
15.....	.5374	.0149	.5474	.0214

It should be borne in mind that the reliability coefficients of .8198 for the miscellaneous group and .8633 for the age-control group are obtained from the results of Series A (1922) correlated with Series B (1923) given to the same children after an interval of one year. The effect of so long a period intervening between examinations is presumably to reduce the coefficient. McCall ('23, page 112), speaking of reliability, states that the best intelligence tests have a self-correlation of from .90 to .95, and that most standard tests have a reliability of about .80. While tests having such a low reliability do not yield sufficiently reliable scores for dealing with individual cases, they give, according to this author, a sufficiently reliable mean score for groups of forty or more. The reliability of the battery, then, meets the present standard, according to this statement.

WEIGHTING OF THE TEST SCORES

Whenever test scores of a given individual are combined into groups of any kind the significance of the total score is determined in part by the balance or weightings of the several test

results of which the group is composed. Weighting should be determined by the relative values of the tests, but these values, in turn, can be determined only when, in addition to their inter-correlations, their correlations with an acceptable criterion are known. A criterion by which to correlate each of the tests and so to determine by the regression method the proportional value of each as it enters into the groups of word, number or spatial tests and, again, as it enters into the groups of analogies, generalization or completion tests or into the composite, is lacking. Such a criterion would be of great value; how soon it will be established is not known. It constitutes a problem, however, which lies outside the scope of the present study, which is concerned with the influence of differences in content and form upon the relative results of tests as they are now used and weighted. While the details of scoring and weighting are too involved to be presented here, the influence of the weightings adopted in the scoring previously described is shown in Table 1.

It is evident that there is rough correspondence between the standard deviations and the time allowed for the respective tests; this relation corresponds fairly closely with a general estimate of the value of the tests. Having used these data in his study of "Mental Discipline in High School Studies," Thorndike ('24, page 6) considers the weightings satisfactory.

CORRELATION METHODS EMPLOYED

Linearity, as required in the use of the Pearson products-moment formula has been assumed throughout. This assumption is somewhat justified first on the basis of the small departure from rectilinearity of the sample tested (Appendix, page 67), and second on the rather generally accepted use of this formula in psychological studies. While the use of the correlation ratio would give equal or higher values for the coefficients, it is assumed that the differences would be too slight to justify the tremendous labor that would be required. To facilitate the computation of so large a number of coefficients as contemplated in this study, the use of a calculating machine seemed a necessity. The adaptation of the Pearson products-moment formula to the operations performed by a calculating machine has been described in full by Toops ('21) and the formulae presented

there have been used throughout. The method is the usual one of computing the standard deviations and mixed moment from a guessed average and making the proper correction for the error. In this adaptation of the formula the plan is to group scores into class intervals not exceeding eighteen in number, and to use transmuted scores, which represent such groupings as deviations from zero as a guessed mean, correction for which is made by the formula. The raw scores have been transmuted in accordance with Tables 23 and 24 (Appendix, pages 65, 66).

CHAPTER III

THE SUBJECTS

THE MAJOR STUDY

Selection of Subjects: The battery of tests was given twice. The first application of the tests was made in May, 1922. School psychologists, members of the research bureau, principals, supervisors, and teachers administered the tests. The subjects were pupils in the ninth, tenth, and eleventh grades. The age of each pupil was obtained at the time of the examination and throughout this study whenever a specific age is given, the age as obtained on the first examination booklet (May, 1922) is that to which reference is made. The second application of the tests, Series B being used, was made one year later (May, 1923) as nearly as possible to the same children, most of whom were at that time in the tenth, eleventh, and twelfth grades. Those who had taken the tests in the ninth grade in 1922 and who had not been promoted to the tenth grade were not given the second examination. Likewise, there were a considerable number who received the second examination who had not been given the first. All such cases have been discarded from this study; only those results have been used which are complete for each of the two examinations given, as described, one year apart.

The school system from which the results have been obtained is one of several selected by the Division of Psychology, Institute of Educational Research, Teachers College, as one in which to conduct its study of "Mental Discipline in High School Studies" and must have been considered representative, although the factors determining its selection apparently have not been published. This school system is in a thriving city in the Middle West, and in Thorndike's ('24 A, page 29) first report is designated as "City I." Table 3 gives the population according to the Federal Census of 1920.

TABLE 3
POPULATION OF "CITY I"
[From Federal Census 1920]

All Classes	Male	Female
Native White.....	162,363	162,048
Foreign White.....	15,377	11,943
Negro.....	15,377	15,247

POPULATION OF HIGH SCHOOL AGES

Classes	Male	Female
Native		
10-14 Years.....	11,084	11,084
15-19 Years.....	11,040	12,865
Native Whites		
10-14 Years.....	9,915	9,852
15-19 Years.....	9,631	11,149
Foreign Born		
10-14 Years.....	350	325
15-19 Years.....	548	590
Negro		
10-14 Years.....	814	876
15-19 Years.....	851	1,123

The distribution of population appears to be approximately that of a fairly representative American city; the population of high school ages studied here is dominantly that of native whites. Moreover, since colored children attend segregated schools and no records from these have been included in this study, there seems to be no adequate reason to question the selection of this population as fairly representative of white pupils of American high schools. The particular schools of this city in which the tests were given are designated in the original description (Thorndike '24 A) as schools 12, 13, 14, and 15. Within these schools approximately five thousand children of the grades named were examined twice in accordance with the original plan, and their records are available. A reduction in the size of the group seemed necessary, for the number of cases is so large that a fairly extensive study of intercorrelations of

test results could not have been completed within a reasonable time. The small improvement in reliability which would have been obtained in a study of so large a population did not seem to justify the undertaking.

Selection of Records: The selection of records was made as follows: In the Division of Psychology, Institute of Educational Research, Teachers College, the scores of these approximately five thousand high school students are recorded on loose cards. The girls' scores are recorded on buff cards; those of the boys on white cards. These original records were examined and disposed of in the order found. To eliminate sex as a factor which might—although there seems little reason to believe that it would—affect the results of the study, the records of boys only have been arbitrarily chosen. All buff cards, records of girls, therefore, were discarded. Scores on the alternate white cards were then recorded. It was estimated that approximately one thousand cases could be selected in this way. This entire group without further selection or elimination was accepted, and has been used in the computations of this study, since it is believed to be fairly representative of a random sampling of boys' records. The writer knows of no reason why it should not be so regarded. When the work was completed, it was found that there were one thousand thirty-nine cases in the group.

Age-Control Group: To have a group in which the factor of age could be rendered more or less constant, a second and overlapping group was simultaneously constructed. This group is composed of two sections: (a) Those in the miscellaneous group of one thousand thirty-nine boys, two hundred fifty-two of whom in the first examination (May, 1922) reported their ages as sixteen years or more, but less than seventeen years; and (b) all other boys, two hundred thirty-seven, within the age limits named, whose records did not fall on the alternate white cards selected, but upon those alternate cards which otherwise would have been discarded. Combining the two groups of boys within the age limits named, there results a group of four hundred eighty-nine cases which will hereafter be referred to as the age-control group of sixteen years.

All computations made from data of the miscellaneous group have been duplicated with parallel computations for the age-control group.

THE TIME-CONTROLLED TEST

Desirability of Control Tests: When these tests of Selective and Relational Thinking and Generalization and Organization were administered, the original directions permitted variation in time spent on the separate tests. Instead of rigidly requiring the subjects to work upon a designated test during the time allotted to it, opportunity was offered them to work ahead as rapidly as possible and to omit any elements they could not do. Upon expiration of the time allowed for the test, however, the subjects were given such directions as: "Even if you have not finished Test 1, begin Test 2,"¹ and so on, but no effort was made to require subjects to comply with this instruction.

Since this study has been made upon the results of tests administered with this method, which will be called *fluid time allowance*, it seemed advisable to determine the relative effects of fixed and fluid time procedures. A fixed time-controlled test was given in which the originally prescribed time intervals were strictly enforced.

Preparation of Papers: The test papers were carefully prepared in advance to assist examiners in insuring rigid time restrictions for each test. Three devices were used. First, each of the fifteen tests was conspicuously numbered with an "Economy Sign Marker" in printer's black ink with figures one inch high. This was done to assist examiners to detect any subject who might turn to some test other than the one upon which the group was working. Second, sheets of paper were fastened between the leaves of the test booklets to cover the tests in such a way that none except the one named by the examiner could be seen. These were fastened with Hotchkiss staples. Finally, a staple was inserted at the loose edge of each booklet so that no page could be exposed without breaking the seal so formed. Test instructions were revised (see Appendix II, pages 68-72) in order to make entirely clear to the subjects how and when the papers were to be opened. All other conditions were kept as nearly as possible the same as for the fluid time procedure.

¹ The quotation is from the original instructions for giving the I. E. R. Tests of Selective and Relational Thinking and of Generalization and Organization. See footnote, page 5.

To try out the arrangement of papers an experimental test was given to a class of forty-four members of the twelfth grade of the high school of Cliffside Park, New Jersey. This test was administered by the writer. During the entire examination not one difficulty was detected. It appeared that the students considered the seal and cover sheets an integral part of the test. With this evidence of practicability, test papers were similarly prepared for the control examination.

Selection of Subjects and Administration of Tests: The subjects selected for the control test were ninth-grade children of a high school having a total enrollment of five hundred fifty. The school is located in New Jersey near New York City. The examinations were given simultaneously in the "home rooms," the groups averaging about thirty-three, with a single group of forty-five, and none larger. There were two hundred thirty-four boys and girls of this grade who appeared to be awaiting for the test eagerly. The seven examiners were graduate students in Teachers College who had previous experience or training in mental testing. Reports of examiners gave no indication of any difficulty with the examination papers as prepared, and suggested no important departure from the standard conditions for giving them. This test, Series A being used, was administered at one o'clock, Thursday, January 10, 1924.

In order to have two sets of scores as in the major group, Series B was given similarly to the same grade on March 7, 1924. These booklets had been numbered, interleaved, and sealed as previously. Graduate students of psychology at Teachers College, Columbia University, again administered the tests. Conditions were not, however, as satisfactory as during the previous examination. The test was given on Friday afternoon at two o'clock, and all students, excepting those tested, had been dismissed before the arrival of the examiners. Some uneasiness and restlessness which may have had a minor effect upon the test results was reported. Two hundred thirteen children were given the second test. Of this number one hundred eighty-four had also taken the first test. The test was administered with time rigidly controlled, as described in the revised directions for giving the tests (Appendix, pages 68 ff.). The results of the one hundred eighty-four cases are reported in Chapter VI.

SUMMARY OF MATERIAL AND METHOD

It has been the purpose of the preceding pages to outline as briefly as possible the nature of the problem, its setting, and the general plan for its study. To determine the influence, if any, of differences in content or form upon test-results, an attempt was made to select a population which might be considered a random sampling of American high school population. For this purpose there were chosen, as subjects, students of the ninth, tenth, and eleventh grades of a city of the Middle West whose population is composed largely of native whites. Inasmuch as colored children attend separate schools and no data from colored schools have been used, these data are those for white children only.

In the further selection of a group which it was hoped would be fairly representative for the purposes of this study, the influence of sex and age was controlled. To accomplish the first, the records of boys only were arbitrarily chosen; for the second result, an empirical control was obtained by forming a group, within the grades named, of all boys sixteen years or more but less than seventeen years.

To reduce the number still further and to make possible a more extensive study of the data, records were chosen by selecting alternate white cards upon which the scores of boys were recorded.

A time-controlled test was given, records on the duplicate series of tests being obtained, to determine whether the original test-directions which prescribe a *fluid time allowance* had any marked effect upon test results.

Throughout, the scores as determined and used in the original study (Thorndike '24 A) were accepted. These data constitute the basis of this study. While a study has been made of the correspondence of results of single tests by sampling, greater reliability has been sought by grouping tests in accordance with the content and form of the tests. The correspondence of results of tests characterized by differences in content will be studied first.

CHAPTER IV

EFFECT OF DIFFERENCES IN CONTENT

I

METHOD OF GROUPING, AND RESULTS

In Chapters II and III are shown the sources of the data from which computations have been made. The first part of this chapter will deal with the effect of differences in content of tests and test groups upon test results. To increase the reliability of the study, the tests were first combined into groups having similar content. Raw scores for single tests have been combined into three groups on the basis of whether each 's (1) a

TABLE 4
COMPARATIVE STANDARD DEVIATIONS AND MEANS OF SINGLE TESTS GROUPED
ACCORDING TO CONTENT
Miscellaneous Group (1039 Cases)

Number Min- utes Allot- ted	Test Num- ber	Group	Series A (1922)		Series B (1923)	
			S.D.	Mean	S.D.	Mean
		Word				
4	2	Absurdities.....	3.6608	8.8402	3.3296	9.3985
5	4	Wylie Opposites.....	6.5466	24.5096	7.1586	27.3941
4	7	Grammatical Analogies.....	4.4038	6.1261	4.6358	7.4754
4	8	Verbal Analogies.....	4.9472	10.5958	4.9904	12.5515
2	9	Moral Judgment.....	3.9410	12.1915	4.3886	12.3455
3	10	Verbal Generalization.....	5.2096	8.7459	5.4914	8.5438
10	13	Trabue Sentence Completion....	8.8737	28.4393	8.8314	31.5895
		Number				
10	1	Arithmetical Problems.....	8.2304	39.2048	8.1256	37.1300
4	5	Number Series Completion.....	4.5950	8.2993	4.6216	10.1222
3	12	Number Generalization.....	3.2196	4.9418	4.5178	8.4171
12	15	Arithmetical Equations Disar- ranged.....	6.9285	14.1266	8.3913	16.7137
		Space				
4	3	Line Arrangements.....	3.2020	3.2613	4.5362	7.6992
4	6	Spatial Analogies.....	4.3988	5.6737	4.7030	12.3956
3	11	Spatial Generalization.....	3.2314	6.1102	3.3314	6.1882
3	14	Cutting Up Surfaces (After Beta Cut-Up).....	2.4667	8.5173	2.2726	9.4701

word-meaning or verbal test, (2) one primarily of numerical relations; or (3) one of spatial or geometrical relations. These groups are spoken of as groups differing in content, and the interrelations of these constitute the first part of this study. Table 4 gives the standard deviations and means of the respective tests as they were combined to make these three groups of tests which differ in respect to content. To facilitate description, these groups are referred to as (1) *words* (2) *numbers*, and (3) *space*.

As stated before (pages 9 and 10), there appears to be rough correspondence between both the importance of the tests and the time allotted to each, and their standard deviations. Since there is no reliable criterion by which to evaluate them as to their validity as representative measures of whatever is tested by a verbal, numerical, or spatial test, and since the present weighting, as it stands, seems fairly reasonable, it is accepted as a working basis. To obtain scores for the respective groups of tests, the raw scores of the several tests have been added, and the sum transmuted into step-values according to the values given in Table 24 (Appendix, page 66). The means and standard deviations, after combination into groups, are given in Table 5.

TABLE 5

STANDARD DEVIATIONS AND MEANS OF TEST-GROUPS DIFFERING IN CONTENT
Miscellaneous Group (1039 Cases)

Group	Series A (1922)		Series B (1923)	
	S.D.	Mean	S.D.	Mean
Words.....	25.7733	103.8636	24.7581	114.7420
Numbers.....	16.2373	63.3144	19.6456	72.2968
Space.....	8.9472	24.0623	9.6244	36.4568

Age-Control Group (489 Cases)

Group	Series A (1922)		Series B (1923)	
	S.D.	Mean	S.D.	Mean
Words.....	25.9677	106.4196	24.9435	116.6877
Numbers.....	17.5118	67.9008	19.1360	75.3928
Space.....	8.5284	24.9196	10.3260	36.7048

From these data intercorrelations and partial correlation coefficients have been computed. These are self-explanatory and are given in Table 6. To assist in interpreting results more quickly and easily the means of the four cross-correlation coefficients and the partial correlations of these are also given. To make possible an examination of the results after the raw coefficients have been corrected for attenuation (Spearman, '04A, pages 88-89), corrected coefficients are presented, together with the values of the coefficients of partial correlation computed from the corrected coefficients. Finally, there is included (Tables 7, 8, 9, and 10) a statement of the coefficients of correlation between single tests which are similar in form but different in content; and also, between tests that differ both in form and in content. In every case the corresponding values for the age-control group are stated.

TABLE 6
COEFFICIENTS OF INTERCORRELATION BETWEEN TESTS GROUPED ON THE BASIS OF
CONTENT

Groups	Miscellaneous Group (1039 Cases)	P.E.	Age-Control Group (489 Cases)	P.E.
Words, 1922—Words, 1923.....	.7660	.0086	.7965	.0152
Words, 1922—Numbers, 1922.....	.5429	.0148	.5843	.0298
Words, 1922—Numbers, 1923.....	.5703	.0141	.5408	.0320
Words, 1923—Numbers, 1922.....	.5516	.0146	.5985	.0290
Words, 1923—Numbers, 1923.....	.5770	.0140	.6167	.0281
Words, 1922—Space, 1922.....	.4641	.0164	.4768	.0349
Words, 1922—Space, 1923.....	.5048	.0156	.5408	.0320
Words, 1923—Space, 1922.....	.4263	.0171	.4790	.0348
Words, 1923—Space, 1923.....	.4519	.0167	.5108	.0334
Numbers, 1922—Numbers, 1923.....	.7084	.0104	.6999	.0231
Numbers, 1922—Space, 1922.....	.4599	.0165	.4919	.0343
Numbers, 1922—Space, 1923.....	.5339	.0150	.5484	.0316
Numbers, 1923—Space, 1922.....	.4017	.0164	.5184	.0331
Numbers, 1923—Space, 1923.....	.5770	.0140	.5701	.0305
Space, 1922—Space, 1923.....	.6081	.0132	.6088	.0258

TABLE 6 (Continued)
MEAN CORRELATIONS. RAW SCORES

Groups	Miscellaneous Group (1039 Cases)	Age-Control Group (489 Cases)
Words—Numbers.....	.5604	.5851
Numbers—Space.....	.5081	.5322
Words—Space.....	.4617	.5012

PARTIAL CORRELATIONS OF MEAN CORRELATION COEFFICIENTS ¹

Group of Tests	Miscellaneous Group (1039 Cases)	Age-Control Group (489 Cases)
Words—Numbers (Space) ¹	.4265	.4346
Words—Space (Numbers).....	.2481	.2765
Numbers—Space (Words).....	.3394	.3405

COEFFICIENTS CORRECTED FOR ATTENUATION ²

Group of Tests	Miscellaneous Group (1039 Cases)	Age-Control Group (489 Cases)
Words—Numbers.....	.7614	.7532
Words—Space.....	.6797	.7309
Numbers—Space.....	.7565	.8168

COEFFICIENTS OF PARTIAL CORRELATION ³

(Coefficients of Zero Order Previously Having Been Corrected for Attenuation)

Group	Miscellaneous Group (1039 Cases)	Age-Control Group (489 Cases)
Words—Numbers (Space).....	.4981	.3960
Words—Space (Numbers).....	.2447	.3043
Numbers—Space (Words).....	.5024	.5932

¹ This should be read: The correlation between the word-meaning tests and the number-relations tests, when the spatial-relations tests are rendered constant, is .4265, the computation of which is based on the means of the four cross-correlations.

² Computed from the formula, $r_{xy} = \frac{\sqrt{r_{x_1y_2} r_{x_2y_1}}}{\sqrt{r_{x_1x_2} r_{y_1y_2}}}$

³ Computed from the formula, $r_{xy.z} = \frac{r_{xy} - r_{xz} r_{yz}}{\sqrt{1 - r_{xz}^2} \sqrt{1 - r_{yz}^2}}$

TABLE 7

INTERCORRELATIONS OF ANALOGIES TESTS—SPATIAL, GRAMMATICAL AND VERBAL
IN CONTENT BUT SIMILAR IN FORM

Test 6 = Spatial Analogies. Test 7 = Grammatical Analogies. Test 8 = Verbal Analogies

Tests	Miscellaneous Group (1039 Cases)		Age-Control Group (489 Cases)	
	<i>r</i>	P.E.	<i>r</i>	P.E.
Grammatical Analogies, 1922—Spatial Analogies, 1922.....	.3176	.0188	.3540	.0267
Grammatical Analogies, 1922—Spatial Analogies, 1923.....	.3306	.0186	.3845	.0260
Grammatical Analogies, 1923—Spatial Analogies, 1922.....	.2968	.0191	.3152	.0274
Grammatical Analogies, 1923—Spatial Analogies, 1923.....	.3762	.0180	.4261	.0250
Mean Correlation.....	.3303		.3699	
Corrected for Attenuation.....	.5839		.8075	
Grammatical Analogies, 1922—Verbal Analogies, 1922.....	.3798	.0179	.3871	.0259
Grammatical Analogies, 1922—Verbal Analogies, 1923.....	.3908	.0177	.3904	.0258
Grammatical Analogies, 1923—Verbal Analogies, 1922.....	.3815	.0179	.3686	.0263
Grammatical Analogies, 1923—Verbal Analogies, 1923.....	.4189	.0245	.4185	.0252
Mean Correlation.....	.3924		.3912	
Corrected for Attenuation.....	.6380		.6011	
Spatial Analogies, 1922—Verbal Analogies, 1922.....	.3292	.0187	.3062	.0277
Spatial Analogies, 1922—Verbal Analogies, 1923.....	.3319	.0186	.2725	.0282
Spatial Analogies, 1923—Verbal Analogies, 1922.....	.3281	.0187	.4040	.0255
Spatial Analogies, 1923—Verbal Analogies, 1923.....	.3712	.0180	.3837	.0260
Mean Raw Correlation.....	.3401		.3413	
Corrected for Attenuation.....	.6430		.7747	

TABLE 7 (Continued)

PARTIAL CORRELATION OF THE MEANS OF THE COEFFICIENTS

Tests	Miscellaneous Group (1039 Cases)	Age-Control Group (489 Cases)
Spatial Analogies—Grammatical Analogies (Verbal Analogies).....	.2276	.2787
Spatial Analogies—Verbal Analogies (Grammatical Analogies).....	.2425	.2345
Grammatical Analogies—Verbal Analogies (Spatial Analogies).....	.3155	.3034

COEFFICIENTS CORRECTED FOR ATTENUATION

Tests	Miscellaneous Group (1039 Cases)	Age-Control Group (489 Cases)
Spatial Analogies—Grammatical Analogies (Verbal Analogies).....	.2691	.6075
Spatial Analogies—Verbal Analogies (Grammatical Analogies).....	.4263	.6030
Grammatical Analogies—Verbal Analogies (Spatial Analogies).....	.4322	.0025

TABLE 8

INTERCORRELATIONS OF GENERALIZATION TESTS, VERBAL AND NUMERICAL IN CONTENT, BUT SIMILAR IN FORM

Test 10 = Verbal Generalization. Test 12 = Numerical Generalization

Tests	Miscellaneous Group (1039 Cases)		Age-Control Group (489 Cases)	
	<i>r</i>	P.E.	<i>r</i>	P.E.
Verbal Generalization, 1922—Numerical Generalization, 1922.....	.1589	.0203	.1900	.0294
Verbal Generalization, 1922—Numerical Generalization, 1923.....	.1716	.0202	.1852	.0294
Verbal Generalization, 1923—Numerical Generalization, 1922.....	.1977	.0201	.2224	.0289
Verbal Generalization, 1923—Numerical Generalization, 1923.....	.3245	.0187	.3659	.0264
Mean Raw Correlation.....	.2131		.2408	
Corrected for Attenuation.....	.5219		.5337	

TABLE 9

INTERCORRELATIONS OF COMPLETION TESTS VERBAL AND NUMERICAL IN CONTENT, BUT SIMILAR IN FORM

Test 5 = Number Series Completion. Test 13 = Trabue Completion

Tests	Miscellaneous Group (1039 Cases)		Age-Control Group (489 Cases)	
	<i>r</i>	P.E.	<i>r</i>	P.E.
Number Series Completion, 1922—Trabue Completion, 1922.....	.1127	.0207	.2145	.0291
Number Series Completion, 1922—Trabue Completion, 1923.....	.1131	.0207	.1689	.0296
Number Series Completion, 1923—Trabue Completion, 1922.....	.1861	.0202	.3156	.0275
Number Series Completion, 1923—Trabue Completion, 1923.....	.1865	.0202	.2748	.0282
Mean Raw Correlation.....	.1495		.2434	
Corrected for Attenuation.....	.2933		.4310	

TABLE 10

CORRELATIONS BETWEEN TESTS DIFFERENT IN FORM AND DIFFERENT IN CONTENT

Test 6 = Spatial Analogies. Test 10 = Otis Verbal Generalization. Test 13 = Trabue Completion

Tests	Miscellaneous Group (1039 Cases)		Age-Control Group (489 Cases)	
	<i>r</i>	P.E.	<i>r</i>	P.E.
Otis Verbal Generalization, 1922—Spatial Analogies, 1923.....	.2560	.0195	.2500	.0286
Otis Verbal Generalization, 1923—Spatial Analogies, 1922.....	.2643	.0195	.3446	.0269
Mean Raw Correlation.....	.2602		.2973	
Corrected for Attenuation.....	.5253		.7535	
Spatial Analogies, 1922—Trabue Completion, 1923.....	.1666	.0203	.2557	.0285
Spatial Analogies, 1923—Trabue Completion, 1922.....	.2850	.0192	.3495	.0268
Mean Raw Correlation.....	.2258		.3026	
Corrected for Attenuation.....	.4425		.7211	

SIGNIFICANCE OF CORRELATION COEFFICIENTS

Correlation is an accepted instrument for the study of the correspondence or the concomitant variation of two sets of measurements. Caution must be exercised, however, in interpreting the results obtained. The meaning of coefficients of correlation is extremely complex and misinterpretation may be made. A special difficulty and danger seems to exist when one attempts to reason from correlation to causation. Thomson ('21, page 355; '23, pages 150-160) has called attention to this difficulty and has shown in his dice-throwing experiments that certain interpretations of intelligence patterns cannot be made from a study of correlation coefficients. A second difficulty seems to exist in the fact that coefficients are expressed numerically, and it is in accordance with habits to conceive of values so expressed as composed of that number of equal units. But a correlation coefficient, numerically expressed as it is, does not indicate a number of equal units of correspondence. The difference between an r of .20 and one of .30 is much less than that between an r of .90 and one of 1.00. This fact suggests that it may be desirable to choose one concrete and definite use of a correlation coefficient by which to evaluate the coefficients as they have been determined.

Such a notion of a correlation coefficient is found in its value when it is used in predicting a measure of one trait from a known measure of a second. The regression equation concisely expresses the use of r in this sense.¹ Accuracy in prediction made by use of the regression equation is dependent upon the closeness of the correlation which exists between the two traits. If the correlation is zero, no improvement over "optimum" chance prediction² can be made, although the individual's score in the second trait is known. In such a case, the "optimum" predicted score for all individuals, when no other facts are known, is the mean score for all individuals possessing that trait; the standard deviation is least when computed from this value (Yule '22, page 135). The predictive value can be determined by comparing the dispersion of real scores about their

$$(X - \bar{X}) = r \frac{\sigma_x}{\sigma_y} (Y - \bar{Y})$$

¹ $\bar{X}$ is the mean of the scores in the trait labeled X .
 $\bar{Y}$ is the mean of the scores in the trait labeled Y .

² See footnote, page 31.

regression values with the dispersion of the measures about the mean score of the trait itself. If correlation is not zero the optimum score for prediction is no longer the mean of the distribution as a whole, but the mean score in the first trait of those having a given score or measure in the second trait. This is graphically represented by the change in position of the regression lines. In zero correlation the lines of best fit to the means of the arrays, or regression lines (Yule '22, pages 175 ff.) as they are called, pass through the means of the distributions at right angles to each other and coincide with the straight lines drawn through these means parallel to the major axes. When correlation increases, the regression lines move toward each other, scissors-like, about the point which represents the means of both traits. Thus they are thrown out of alignment with the axes drawn through the means and the variability of the scores of each array about their own regression values is less than the variability of the trait as a whole. The reduced variability is given by the equation,

$$\sigma_{1.2} = \sigma_1 \sqrt{1 - r^2_{12}}$$

which is the standard error of a score predicted by the use of the regression equation (Kelley, '23, pages 160-161, and Yule, '22, pages 175-180) and is called the standard error of estimate. If both sides of the equation are divided by σ_1 and each side is subtracted from unity, the formula obtained is:

$$\frac{\sigma_1 - \sigma_{1.2}}{\sigma_1} = 1 - \sqrt{1 - r^2_{12}}.$$

The Predictive Index: The numerator of the fraction on the left side of the equation is clearly the amount by which error in prediction has been reduced, and the ratio of this difference to the standard deviation is a measure of the predictive value of the correlation coefficient and will be called the *predictive index* (P.I.).

$$\begin{aligned} \text{P.I.} &= 1 - \sqrt{1 - r^2_{12}} \\ (\text{P.I.} - 1)^2 + r^2_{12} &= 1 \end{aligned}$$

This is the equation of a circle. The upper two quadrants of it have been discarded (see Figure 1) because $\sigma_{1.2}$ cannot be less than zero nor greater than σ_1 . From this it clearly follows

that the values of the predictive index (P.I.) range from 0 to +1. Figure 1 graphically presents the relative predictive values for different correlation coefficients. Table II shows the numerical values of certain coefficients of correlation expressed in terms of their respective predictive indices.

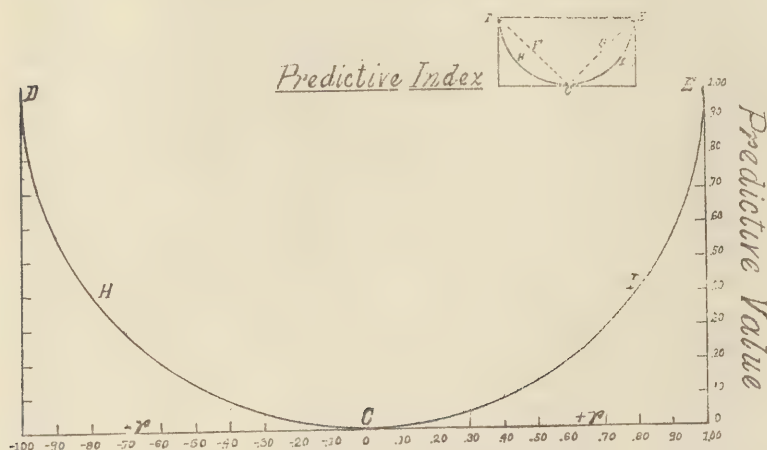


FIG. 1. The values of the correlation coefficient ranging from -1 through 0 to $+1$ are plotted horizontally; the predictive values which range from 0 to $+1$ are plotted on the vertical scale.

Example of Use. To determine the approximate predictive index for a correlation coefficient $.80$ proceed as follows: Locate $.80$ on the horizontal scale. Determine the point at which the ordinate erected at $.80$ is crossed by the arc *DHCIE*. Read the value of this point on the vertical scale at the right. In this case it is $.40$, which is the predictive index for a coefficient of correlation of $.80$.

The insert suggests the difference between a straight line concept of the predictive values of correlation coefficients shown by the dotted lines *DFC* and *CGE* and the true predictive values shown by the arc *DHCIE*.

The predictive index is a number which expresses the amount of improvement in prediction which is made possible by reason of the correlation which exists between the two traits considered. It is therefore a measure of the significance of the correlation coefficient from the predictive viewpoint, and will be used here, in the sense described, in evaluating the coefficients which have been determined in this study.

The effect of so-called errors in measurement upon the magnitude of correlation coefficients should be briefly considered. Variation in psychological measurement is a reality of such importance that its ever-present occurrence and effect must be

TABLE 11
PREDICTIVE INDICES FOR CERTAIN CORRELATION COEFFICIENTS

r	P.I.	r	P.I.	r	P.I.
.00	.0000				
.01	.0001	.36	.0671	.71	.2958
.02	.0002	.37	.0710	.72	.3061
.03	.0005	.38	.0751	.73	.3166
.04	.0008	.39	.0792	.74	.3274
.05	.0013	.40	.0835	.75	.3386
.06	.0019	.41	.0880	.76	.3501
.07	.0025	.42	.0925	.77	.3620
.08	.0033	.43	.0972	.78	.3743
.09	.0041	.44	.1021	.79	.3869
.10	.0051	.45	.1070	.80	.4000
.11	.0061	.46	.1121	.81	.4136
.12	.0073	.47	.1174	.82	.4277
.13	.0085	.48	.1228	.83	.4423
.14	.0099	.49	.1283	.84	.4575
.15	.0114	.50	.1340	.85	.4733
.16	.0129	.51	.1399	.86	.4998
.17	.0146	.52	.1459	.87	.5070
.18	.0164	.53	.1521	.88	.5251
.19	.0182	.54	.1584	.89	.5441
.20	.0203	.55	.1649	.90	.5642
.21	.0222	.56	.1716	.91	.5854
.22	.0245	.57	.1794	.92	.6081
.23	.0269	.58	.1854	.93	.6325
.24	.0293	.59	.1926	.94	.6589
.25	.0318	.60	.2000	.95	.6878
.26	.0343	.61	.2076	.96	.7200
.27	.0372	.62	.2154	.97	.7569
.28	.0400	.63	.2235	.98	.8011
.29	.0420	.64	.2317	.99	.8590
.30	.0461	.65	.2401	1.00	1.0000
.31	.0493	.66	.2488		
.32	.0526	.67	.2577		
.33	.0561	.68	.2668		
.34	.0596	.69	.2762		
.35	.0633	.70	.2859		

recognized. Among the factors which contribute to these so-called errors in measurement are the facts that tests are imperfect instruments of measurement and that there is chance variability of individuals' scores, and probably other causes,

as well as literal errors of measurement.¹ All such random factors tend to reduce the coefficients below their theoretically true values. Correction of coefficients for attenuation (Spearman, '04A; Yule, '22, page 313) is statistical compensation for such errors of measurement. If the assumptions made in the derivation of the formula for this correction are true in the particular cases to which the formula is applied, the values so determined are theoretically the true values. The formula requires approximate linearity and demands that errors of measurement be uncorrelated (*a*) with each other and (*b*) with the measures themselves. In actual practice it is probable that these conditions are seldom realized. Because of these facts, it is altogether probable that the coefficients as corrected are too high. Tests of the assumptions involved have not been made; they should have been. For theoretical study, then, properly corrected coefficients may be regarded as the truest obtainable measures of linear correspondence of traits or abilities (Spearman, '04A). The corrected coefficients used in this study have been computed in accordance with the short formula² because of the probable reduction in correlation of errors. The relative values and significance of corrected coefficients between tests which differ in content will be examined first.

II

CORRESPONDENCE OF RESULTS OF TESTS HAVING DIFFERENCES IN CONTENT

Test Groups: For the miscellaneous group of one thousand thirty-nine cases the corrected coefficient of correlation between words and numbers (Table 6) is .7614; between numbers and space, .7565; and between words and space, .6797. The difference between the first two is small and is not regarded as significant. Coefficients such as these have been regarded as

¹ There may be some actual errors due to administration, writing, scoring, tabulating, and computing the results presented here. Some of these steps had been completed before the present study was begun. Since the tests were given and scored under the direction and supervision of the Institute of Educational Research, Division of Psychology, Teachers College, it is believed that errors have been minimized, at least to the extent that the results are fairly representative of the best results of tests as commonly administered and scored. Care has been taken to prevent the occurrence of errors in the operations performed on a calculating machine by generally repeating such operations. In spite of the care that has been taken throughout in checking, transcribing, tabulating and computing, it is entirely possible that some errors may be found in the present work.

²
$$r_{xy} = \frac{\sqrt{r_{x_1y_2} r_{x_2y_1}}}{\sqrt{r_{x_1x_2} r_{y_1y_2}}}$$
 Spearman '04A; also, Yule '22, page 213.

high (Rugg, '17, page 258), but when their values for prediction are seen to be respectively .3671, .3511, and .2646, their wide departure from perfection is clearly evident. Or, in other words, if scores in the test in numbers should be predicted from the corresponding scores in words, .3671, or less than three-eighths, of the error of optimum chance prediction (see page 31) has been eliminated but .6329, or more than five-eighths of it, still remains. For estimating a score in spatial tests from a known score in number tests, the predictive value is approximately equal to that just stated. Even less predictive value is found in the correlation between words and space as seen in its predictive index, .2646. Variability of scores from predicted values in this case, i.e., of a score in words from one in space, or vice-versa, is reduced by slightly more than one-fourth of the variability in scores of the predicted trait, but nearly three-fourths still remains!

Further examination of these coefficients is interesting. When ability as measured by the spatial tests has been rendered constant by partial correlation, the remaining correspondence between the verbal and numerical tests is .4981; with words rendered constant, the correlation between space and numbers is .5024; and with numbers rendered constant, the correlation between the spatial and verbal tests is .2447. The predictive indices, or measures of goodness of these coefficients, are .1329, .1354, and .0304.

The difference between these values and the values of their corresponding predictive indices for the r 's of the zero order, gives a statement of the amount of predictive value which has been lost by eliminating the influence of the traits which have been rendered constant. In the first case, when space is rendered constant, the predictive value of the r between words and numbers is reduced from .3671 to .1329, or .2342. There appears, therefore, to be some trait or ability measured by the spatial test-group which is common to the word or number test-groups. However, there still remains a correlation of .4981, the predictive index of which is only .1329. A very similar statement could be made in the second case by properly interchanging the names of the test groups. In the third case, the largest reduction in value is seen; by rendering numbers constant the correlation between words and space is reduced from

.6797 to .2447, the predictive index of which is .0304. Only slight improvement over chance¹ prediction of a trait remains. It is interesting to note that the predictive values which have been subtracted in each case are approximately equal. They are .2342, .2217, and .2342, for the three respective cases. It should be borne in mind that these are corrected coefficients and are considered the truest and highest obtainable measures of correspondence between the traits tested. When considered from this point of view, it is clearly seen how unreliable it is to forecast performance in any of these tests from results in another, although in every case this is better than a chance prediction.

Passing on to an examination of the uncorrected, or raw coefficients of correlation (Table 6), it is seen that there is marked constancy in the values of the four corresponding cross-coefficients, and the complexity of treating each separately would be great. It appears advisable to confine this discussion to the means of these coefficients. The mean correlation between words and numbers is .5604; between numbers and space, .5081; and between words and space, .4617. It is seen that the values fall in the same order as found in the corrected coefficients. The predictive indices in the same order are .1718, .1307 and .1130. When the formula for partial correlation is used, the third variate in each case having been rendered constant, the correlations are: between words and numbers, .4265; between words and space, .2481; and between numbers and space, .3394. This order is different from that of the coefficients of partial correlation of corrected coefficients. The coefficients of correlation between words and space with numbers rendered constant are practically identical in the two cases. The predictive indices for these coefficients are, respectively, .0955, .0313, and .0594, and indicate the effective predictive value of these coefficients.

The partial correlation coefficients between each pair of analogies, the third in each case having been rendered constant, are: spatial and grammatical, .2276; spatial and verbal, .2425;

¹ Chance prediction does not mean a wild guess, irrespective of the mean or variability of the trait in question nor even the random drawing of a single score from all the scores of the trait; there is meant, rather, the optimum predicted score, which, in a frequency distribution of a single trait and when no other facts are known, is the *mean* of the trait from which point the root mean square deviation is least.

grammatical and verbal, .3155. Regarding the order of magnitude, they follow the order which might have been guessed. It would be expected that the highest correlation would be found between the grammatical and verbal analogies since both are verbal in content. The other coefficients vary little, but it seems reasonable to believe that the grammatical test measures a narrower function than the verbal test, and therefore might be expected to correlate less with a test of very different content than with the test of wider functions. There is, however, less difference than might have been guessed. The predictive indices for these coefficients are .0262, .0298, and .0511.

DIFFERENCE BETWEEN SINGLE TESTS

The correlations just described have been computed between groups of tests of verbal, numerical, and spatial content. As such, the forms of tests in the various groups so constituted are different, and it appears that this is a factor whose influence should be controlled as far as possible. For this purpose correspondence between single tests having the same form, but different content, have been computed. These will be studied under four headings: (a) spatial, grammatical,¹ and verbal tests, all of the analogies form; (b) tests of verbal and numerical content, but both of the generalization or similarities form; (c) number series completion and sentence completion tests; and (d) correlations between tests different in both content and form, in which the correlation of the geometrical or spatial analogies test with the sentence completion test and the verbal generalization test will be studied.

a. Tests of the Analogies Form: The intercorrelations of the tests of spatial, grammatical, and verbal content, all of the analogies form, are given first in Table 7. The form of the tests has been rendered constant empirically, at least as nearly as it seems possible to make tests of the same form but of different content. With the effect of form rendered constant in this way, the coefficients of intercorrelation between the spatial, grammatical, and verbal tests are nearly equal. The corrected coefficients of correlation between the grammatical analogies and the spatial analogies is .5839; between the verbal analogies and the grammatical analogies .6380; and between the spatial analo-

¹ Used for the study, although it is verbal and not numerical in content.

gies and the verbal analogies, .6430. It is interesting to notice that the first named is the lowest, the last named the highest; but the differences are small. The predictive indices are .1882, .2300, and .2341, respectively.

The coefficients of partial correlation for corrected coefficients between each two, when the third, in turn, has been rendered constant, are in the case of spatial or geometrical analogies and grammatical analogies, .2691 (P.I. .0369); between spatial or geometrical analogies and verbal analogies, .4263 (P.I. .0954); and lastly, between grammatical analogies and verbal analogies (spatial analogies having been rendered constant), .4322 (P.I. .0992). The expectation would have been reasonable that the correlation between the first two, spatial and grammatical analogies, would have been the smallest, and that grammatical and verbal analogies would have correlated the highest. The difference between the latter and the correlation between spatial analogies is too slight to be considered very significant, as is shown by the predictive indices.

The same trend of the results exists in the uncorrected coefficients. Their mean values, .3303, .3924, and .3401, fall in the same order as the corrected coefficients and their variation is not great. They are interesting chiefly because of the persistency of positive correlation, however different the content is, and, on the other hand, because the coefficients are not higher although the same form of test is used throughout.

b. Tests of the Generalization Form: The next study made is of the correspondence between a verbal test and a numerical test when both are of the generalization form (Table 8). The corrected coefficient of this correlation is .5219 (P.I. .1470) while the mean of the intercorrelations is .2131 (P.I. .0230). The low positive correlation found in the preceding section persists, although both coefficients are even lower than between the tests in which the analogies form was maintained throughout.

c. Tests of the Completion Form: In the case where the completion form of tests is used and the tests are number series and sentence completion, the coefficients are even lower (Table 9). The corrected coefficient is .2933 (P.I. .0440) and the mean of the raw cross coefficients is .1495 (P.I. .0112).

d. Tests Different in Both Form and Content: Finally, Table 10 shows the correlation between two pairs of tests which differ

both in content and form according to classification used in this study. The corrected correlation coefficient between the spatial or geometrical analogies, Test 6, and the sentence completion, Test 13, is .5253 (P.I. .1491) while the mean of the raw coefficients is .2602 (P.I. .0344). Between the spatial analogies test and the verbal generalization test, the corrected coefficient as determined is .4425 (P.I. .0133) and the mean of the raw coefficients is .2258 (P.I. .0259). These values are higher than that between the two tests in which the completion form was kept constant; however, by inspection it can be seen that they are slightly lower than the mean correlation between single tests.

CHAPTER V

EFFECT OF DIFFERENCES IN FORM

I

METHOD OF GROUPING AND RESULTS

The second phase of the analysis of the results is made in the same manner as the first. This phase has to do with the correlation of tests and groups of tests which are different in form, when by differences in form is meant those differences which exist between completion tests, analogies tests, and generalization tests. The selection of tests found in the battery

TABLE 12

COMPARATIVE STANDARD DEVIATIONS AND MEANS OF SINGLE TESTS GROUPED
ACCORDING TO FORM

Miscellaneous Group (1039 Cases)

Number Min- utes Allot- ted	Test Num- ber	Group	Series A (1922)		Series B (1923)	
			S.D.	Mean	S.D.	Mean
		Completion				
4	5	Number Series Completion.....	4.5950	8.2993	4.6216	10.1222
10	13	Trabue Sentence Completion....	8.8737	28.4393	8.8314	31.5895
		Analogies				
4	6	Spatial Analogies.....	4.3988	5.6737	4.7030	12.3956
4	7	Grammatical Analogies.....	4.4038	6.1261	4.6358	7.4754
4	8	Verbal Analogies.....	4.9472	10.5958	4.9904	12.5515
		Generalization				
3	10	Verbal Generalization.....	5.2996	8.7459	5.4914	8.5438
3	11	Spatial Generalization.....	3.2314	6.1102	3.3314	6.1882
3	12	Number Generalization.....	3.2196	4.9418	4.5178	8.4171
		Composite ¹				
10	1	Arithmetical Problems.....	8.2304	39.2048	8.1256	37.1300
4	2	Absurdities.....	3.6608	8.8402	3.3296	9.3985
4	3	Line Arrangements.....	3.2020	3.2613	4.5362	7.6992
5	4	Wylie Opposites.....	6.5466	24.5096	7.1586	27.3941
2	9	Moral Judgment.....	3.9410	12.1915	4.3886	12.3455
12	15	Arithmetical Equations Disar- ranged.....	6.9285	14.1266	8.3913	16.7137

¹Described in Section II, page 43.

of the Selective and Relational Thinking and Generalization and Organization Tests makes possible also a comparison of single tests of certain unusual similarities. These will be studied; but first, in order to have data upon which more reliable computations can be made, tests of similar form have been grouped as shown in Table 12, which gives also their comparative standard deviations and means. The means and standard deviations of the groups, after the tests have been combined by the addition of the respective raw scores, are given in Table 13.

TABLE 13
STANDARD DEVIATIONS AND MEANS OF TEST-GROUPS DIFFERING IN FORM
Miscellaneous Group (1039 Cases)

Group	Series A (1922)		Series B (1923)	
	S.D.	Mean	S.D.	Mean
Completion.....	10.6542	38.2170	11.0277	43.3470
Analogies.....	10.3827	26.4042	10.8006	36.4612
Generalization.....	8.3673	20.7878	10.1809	24.3681
Composite.....	20.7751	100.9480	23.3435	109.4956

Age-Control Group (489 Cases)

Group	Series A (1922)		Series B (1923)	
	S.D.	Mean	S.D.	Mean
Completion.....	11.5856	38.9215	11.6065	43.5122
Analogies.....	10.2887	26.2372	11.1798	36.2740
Generalization.....	8.4253	20.5859	10.5752	24.2669
Composite.....	22.0938	103.4867	23.0829	111.8507

CORRESPONDENCE OF RESULTS OF TEST-GROUPS HAVING DIFFERENCES IN FORM

As in the previous case when tests were combined to make similar content-groups (page 19), the weightings of the component tests, as estimated from their standard deviations, were accepted as satisfactory. Any error involved in such a decision has not been corrected, and computations have been based upon the raw scores so weighted. After an examination of their inter-correlations, a comparison of each group of tests with a com-

posite¹ made up of six tests, none of which is included in the groups based on the form of test, will be made. As before, corrected coefficients will first be studied upon the assumption that they represent theoretically, with the effect of errors² eliminated as far as possible, the truest obtainable measure of correspondence. Later, for practical considerations, the raw coefficients will be studied. The coefficients of intercorrelation are presented in Table 14.

TABLE 14

COEFFICIENTS OF INTERCORRELATION BETWEEN TESTS GROUPED ON THE BASIS OF FORM

Groups of Tests	Miscellaneous Group (1039 Cases)	P.E.	Age-Control Group (489 Cases)	P.E.
Composite, 1922—Composite, 1923.	.7132	.0103	.7552	.0131
Composite, 1922—Analogies, 1922.	.6164	.0130	.6795	.0164
Composite, 1922—Analogies, 1923.	.6270	.0127	.6782	.0164
Composite, 1923—Analogies, 1922.	.5999	.0134	.6737	.0166
Composite, 1923—Analogies, 1923.	.6281	.0127	.6884	.0160
Composite, 1922—Generalization, 1922.	.4552	.0166	.4681	.0238
Composite, 1922—Generalization, 1923.	.4551	.0166	.4757	.0236
Composite, 1923—Generalization, 1922.	.4214	.0172	.4276	.0249
Composite, 1923—Generalization, 1923.	.4296	.0105	.5048	.0227
Composite, 1922—Completion, 1922.	.5449	.0147	.6143	.0186
Composite, 1922—Completion, 1923.	.5596	.0144	.5923	.0198
Composite, 1923—Completion, 1922.	.5207	.0153	.5634	.0208
Composite, 1923—Completion, 1923.	.5678	.0142	.6684	.0167
Analogies, 1922—Analogies, 1923.	.7193	.0101	.7421	.0137
Analogies, 1922—Generalization, 1922.	.4938	.0158	.4932	.0231
Analogies, 1922—Generalization, 1923.	.4853	.0160	.5073	.0226
Analogies, 1923—Generalization, 1922.	.4471	.0167	.4492	.0243
Analogies, 1923—Generalization, 1923.	.4873	.0160	.4926	.0231
Analogies, 1922—Completion, 1922.	.4729	.0162	.5403	.0216
Analogies, 1922—Completion, 1923.	.5250	.0152	.5136	.0224
Analogies, 1923—Completion, 1922.	.4999	.0157	.5560	.0211
Analogies, 1923—Completion, 1923.	.5252	.0151	.5313	.0219
Generalization, 1922—Generalization, 1923.	.5643	.0143	.5915	.0198
Generalization, 1922—Completion, 1922.	.3090	.0190	.3248	.0273
Generalization, 1922—Completion, 1923.	.3616	.0182	.3209	.0274
Generalization, 1923—Completion, 1922.	.3276	.0187	.3522	.0267
Generalization, 1923—Completion, 1923.	.3411	.0185	.3328	.0271
Completion, 1922—Completion, 1923.	.5440	.0147	.6071	.0193

¹ Described in Section II, page 43.

² See p. 27.

TABLE 14 (Continued)
COEFFICIENTS CORRECTED FOR ATTENUATION

Groups of Tests	Miscellaneous Group (1039 Cases)	Age-Control Group (489 Cases)
Analogies—Completion.....	.8190	.7961
Analogies—Generalization.....	.7311	.7205
Completion—Generalization.....	.6212	.5610
Composite—Analogies.....	.8563	.9029
Composite—Generalization.....	.6003	.6748
Composite—Completion.....	.8666	.8531

COEFFICIENTS OF PARTIAL CORRELATION

Groups of Tests	Miscellaneous Group (1039 Cases)	Age-Control Group (489 Cases)
Composite—Completion (Generalization).....	.7722	.8540
Composite—Analogies (Generalization).....	.7123	.8143
Analogies—Completion (Generalization).....	.6823	.6827
Completion—Composite (Analogies).....	.5576	.5162
Analogies—Composite (Completion).....	.5116	.7086
Analogies—Generalization (Completion).....	.4945	.5467
Generalization—Composite (Completion).....	.3887	.4542
Generalization—Composite (Analogies).....	.1823	.0813
Generalization—Completion (Analogies).....	.0573	.0300
Completion—Composite (Analogies—Generalization)...	.5573	.5206
Analogies—Composite (Completion—Generalization)...	.3957	.6169
Generalization—Composite (Analogies—Completion)...	.1814	.1131

MEAN COEFFICIENTS OF CORRELATIONS BETWEEN GROUPS CLASSIFIED
ON THE BASIS OF FORM

Raw Data

Groups of Tests	Miscellaneous Group (1039 Cases)	Age-Control Group (489 Cases)
Composite—Analogies.....	.6179	.6799
Composite—Completion.....	.5483	.6000
Analogies—Completion.....	.5058	.5353
Analogies—Generalization.....	.4784	.4855
Composite—Generalization.....	.4403	.4698
Completion—Generalization.....	.3348	.3326

TABLE 14 (Continued)
 PARTIAL CORRELATIONS OF MEANS OF COEFFICIENTS

Groups of Tests	Miscellaneous Group (1039 Cases)	Age-Control Group (489 Cases)
Analogies—Generalization (Completion).....	.3802	.3860
Analogies—Completion (Generalization).....	.4177	.4534
Completion—Generalization (Analogies)	.1225	.0985
Composite—Analogies (Completion).....	.4721	.5281
Composite—Analogies (Generalization).....	.5165	.5854
Composite—Generalization (Completion).....	.3257	.3572
Composite—Generalization (Analogies).....	.2096	.2179
Composite—Completion (Analogies).....	.3476	.3966
Composite—Completion (Generalization).....	.4734	.5445
Composite—Analogies (Completion—General).....	.3983	.4548
Composite—Completion (Analogies—General).....	.3317	.3863
Composite—Generalization (Analogies—Completion)...	.1795	.1958

Results: For the miscellaneous group of one thousand thirty-nine cases, the analogies group correlates with the completion group, .8190; the analogies group with the generalization group, .7311; and the generalization group with the completion group, .6212. The predictive indices of these coefficients are .4262, .3177, and .2163. The coefficients of partial correlation between each of the three pairs when the third has been rendered constant are: between analogies and completion, .6823 (P.I. .2689); between analogies and generalization, .4945 (P. I. .1307); and between generalization and completion, .0573 (P.I. .0016). Thus the order of relative magnitude is changed. The correlation between analogies and completion is reduced from .8190 to .6823, but still remains highest; between analogies and generalization, the elimination of the effect of the completion reduced the coefficient by .2366; while the rendering of analogies constant reduced the correlation between generalization and completion by .5639 or almost to zero. Its remaining value of .0573 has a predictive index of .0016.

Examination of the raw data shows the same trend. To escape confusion in trying to examine the many separate cross-correlations, the variations of which appear small, the means of these coefficients are again used. Between analogies and completion (Table 14) the mean correlation is .5058 (P.I. .1373);

between analogies and generalization, .4784 (P.I. .1219); and between completion and generalization, .3348 (P.I. .0577). The values follow the order of the corrected coefficients. The coefficients of partial correlation for these means when the third variable is rendered constant in each case are: analogies and completion, .4177 (P.I. .0914); analogies and generalization, .3802 (P.I. .0751); and completion and generalization, .1225 (P.I. .0075). Again, the order of the corrected coefficients is found to be the same as that of the uncorrected ones. Also the tendency to extreme reduction, when analogies are rendered constant, of the correlation between completion and generalization groups is confirmed.

Table 15 gives the correlation between tests of different form but all of numerical content; Table 16 gives the correlation between two word-meaning tests of different form, and Table 17 gives the correlation between a spatial-construction test and a spatial-analogies test. Examination of these disclose the correspondence between single tests of different form when the content is rendered constant empirically.

a. *Numerical Tests Having Differences in Form:* The correlation between arithmetical problems and number series completion tests, when corrected for attenuation, is .6239 (P. I. .2185); and between number series completion and disarranged arithmetical equations, .5989 (P.I. .1192). The corresponding raw coefficients are .3367 (P.I. .0584) and .3106 (P.I. .1192), respectively. It is clearly evident that these values are lower than for the tests when grouped, but this might be expected because of their lower variability. (See Table 15).

b. *Verbal Tests Having Differences in Form:* Between two tests of different forms which have verbal content, the verbal analogies and the sentence completion test, the corrected correlation is .5097 (P.I. .1396), and the mean raw coefficient is .2701 (P.I. .0372). In this case the correspondence between the tests is slightly smaller than between the numerical tests of different form. It is impossible with the data at hand to tell whether the difference is due to the difference between the forms of the tests or to the differences between the specific items of content. It is plausible to guess that the differences of form are probably more effective than the small differences of content, but there is no proof for the inference. (See Table 16).

TABLE 15

INTERCORRELATIONS OF NUMERICAL TESTS—TESTS OF DIFFERENT FORM
BUT ALL OF NUMERICAL CONTENT

*Test 5 = Number Series Completion. Test 1 = Arithmetical Problems. Test 15 =
Disarranged Arithmetical Equations.*

Tests	Miscellaneous Group (1039 Cases)	P.E.	Age-Control Group (489 Cases)	P.E.
Arithmetical Problems, 1922 — Number Series Completion, 1922.....	.2643	.0195	.3717	.0263
Arithmetical Problems, 1922 — Number Series Completion, 1923.....	.3192	.0188	.0865	.0303
Arithmetical Problems, 1923 — Number Series Completion, 1922.....	.3436	.0185	.4658	.0239
Arithmetical Problems, 1923 — Number Series Completion, 1923.....	.4199	.0172	.1537	.0298
Mean Raw Correlation.....	.3367		.2260	
Corrected for Attenuation.....	.6239		.6465	
Arithmetical Equation, 1922 — Arith- metical Problems, 1922.....	.2175	.0199	.2657	.0283
Arithmetical Equation, 1922 — Arith- metical Problems, 1923.....	.3250	.0187	.3185	.0274
Arithmetical Equation, 1923 — Arith- metical Problems, 1922.....	.2994	.0190	.3136	.0275
Arithmetical Equation, 1923 — Arith- metical Problems, 1923.....	.4006	.0176	.4256	.0250
Mean Raw Correlation.....	.3106		.4307	
Corrected for Attenuation.....	.5989		.5986	

TABLE 16

CORRELATION BETWEEN TESTS OF DIFFERENT FORM BUT BOTH OF
VERBAL CONTENT

Test 8 = Verbal Analogies. Test 13 = Trabue Completion

Tests	Miscellaneous Group (1039 Cases)		Age-Control Group (489 Cases)	
	r	P.E.	r	P.E.
Verbal Analogies, 1922—Trabue Completion, 1922.....	.2811	.0192	.3567	.0266
Verbal Analogies, 1922—Trabue Completion, 1923.....	.2576	.0195	.3010	.0277
Verbal Analogies, 1923—Trabue Completion, 1922.....	.2867	.0192	.3545	.0266
Verbal Analogies, 1923—Trabue Completion, 1923.....	.2554	.0195	.3021	.0277
Mean Raw Correlation.....	.2701		.3286	
Corrected for Attenuation.....	.5097		.5512	

TABLE 17
CORRELATION BETWEEN TESTS OF DIFFERENT FORM BUT BOTH OF
SPATIAL CONTENT

Test 3 = Line Arrangement. Test 6 = Spatial Analogies

Tests	Miscellaneous Group (1039 Cases)	P.E.	Age-Control Group (489 Cases)	P.E.
Line Arrangement, 1922—Spatial Analogies, 1922.....	.2189	.0199	.2366	.0288
Line Arrangement, 1922—Spatial Analogies, 1923.....	.2695	.0194	.2470	.0286
Line Arrangement, 1923—Spatial Analogies, 1922.....	.2692	.0194	.2750	.0282
Line Arrangement, 1923—Spatial Analogies, 1923.....	.2916	.0192	.3172	.0274
Mean Raw Correlation.....	.2622		.2690	
Corrected for Attenuation.....	.5610		.8118	

c. Spatial Tests Having Differences in Form: An analogous case is found in the correlation of the spatial construction test with the spatial analogies test. The corrected coefficient is .5610 (P.I. .1722) and the raw coefficient is .2622 (P.I. .0350).

Table 10 is also interesting here, for it gives, between two tests which are different in form and different in content, spatial analogies, Test 6, and Otis verbal generalization, Test 10, a corrected coefficient of .5253 (P.I. .1460) whose corresponding raw coefficient is .2602 (P.I. .0344). Between spatial analogies, Test 6, and the Trabue completion, Test 13, the corrected coefficient is .4425 (P.I. .1032) and the uncorrected coefficient is .2258 (P.I. .0258). (See Table 17).

II

CORRELATION OF TEST GROUPS WITH A COMPOSITE

It seemed advisable to try to compare, in terms of some criterion, the relative general value of each of the form groups of tests. No adequate criterion was available, for not even scores of general intelligence, other than that of the battery itself, were available for subjects of this group. There were, however, seven of the fifteen tests which did not fall naturally into the groups of analogies, generalization, or completion groups. Six of these were grouped into a composite which promised to be fairly, if not entirely, satisfactory for this purpose. Test 14,

cutting-up surfaces, adapted from Army Beta, was omitted because, when the results of this test were plotted and the distribution was found to be irregular and its standard deviation small, it was felt that its inclusion would add little to the value of the composite. The reasons are probably inadequate to justify its elimination, but computation of results had been made before judgment in this matter was reversed.

As the composite stands, it is made up of Tests 1, 2, 3, 4, 9, and 15. These are tests in arithmetical problems, absurdities, line arrangements, opposites, moral judgments, and disarranged arithmetical equations. For the group of one thousand thirty-nine miscellaneous cases, the composite of Series A, 1922, has a mean of 103.4867 and a standard deviation of 22.0938; and the composite of Series B, 1923, a mean of 111.8507 and a standard deviation of 23.0829. Its reliability coefficient is .7132 with a probable error of .0103.

The problem is this: If only one of these test-groups could be used, which one would give results most closely approximating the results of a composite? To answer the question, correlations between the composite, as a criterion, and the completion, analogies, and generalization tests were computed as shown in Table 14.

It appears that both the completion group and the analogies group correlate, correction having been made for attenuation, rather high and almost equally as well with the composite; the generalization group and composite fail to correlate as high. The coefficient shows that the completion test correlates with the composite the highest, .8666 (P.I. .5010). It is interesting to stop at this point to realize that, for the first time in this study, a correlation coefficient has been obtained which has sufficient value in prediction to reduce the error in that operation by one half or more of the original variability. The fifty per cent mark in this case is just barely passed. Analogies correlates next high with the composite and almost as well, .8563 (P.I. .4835); and generalization correlates least well, with a corrected coefficient of .6903 (P.I. .2765) which is only a little better than one half the predictive value of the coefficient first named in this section.

Among the raw coefficients the order is different, although the variation is not great. Analogies correlates with the com-

posite .6179 (P.I. .2137); the completion with the composite .5483 (P.I. .1637); and generalization with the composite .4403 (P.I. .1022). Here the predictive value of the generalization test is approximately five-eighths of the value of the completion test and just less than one-half that of the analogies test. The apparently low value of this type of test for general ability is confirmed. This tendency is seen again in a study of the coefficient of partial correlation.

When the corrected coefficients are substituted in the formula for partial correlation ¹ the following results are obtained:

	<i>r</i>	P.I
Composite and Completion (Generalization)*	.7722	.3646
Composite and Analogies (Generalization)	.7123	.2991
Composite and Completion (Analogies)	.5576	.1699
Composite and Analogies (Completion)	.5116	.1408
Composite and Generalization (Completion)	.3887	.0786
Composite and Generalization (Analogies)	.1814	.0166

*The first is to be read, "The correlation between the composite and completion, when generalizations have been rendered constant, is .7722." These computations have been made from coefficients of the zero order corrected for attenuation.

When freed from the influence of generalization tests the completion group and composite correlate highest, and analogies stands next in value. There is comparatively little influence exerted on these coefficients by the generalization group. Similarly, the composite and generalization group correlates least when the influence of the completion or analogies group is rendered constant. These coefficients drop, respectively, to .3886 and .1823.

In the coefficients of the second order, interesting differences in rank order appear. The correlation between completion and composite, with analogies and generalization rendered constant, is .5573 (P.I. .1697); between analogies and composite, when completion and generalization are rendered constant, .3957 (P.I. .0816); and between generalization and composite, when analogies and completion have been rendered constant, the smallest correlation of all, is .1814 (P.I. .0166).

The order of the partial raw coefficients varies slightly from this. Here it is seen that the analogies have higher correlation

$$^1 r_{xy.z} = \frac{r_{xy} - r_{xz} r_{yz}}{\sqrt{1 - r_{xz}^2} \sqrt{1 - r_{yz}^2}}$$

with the composite than the completions have, but the difference is not great. The correlation between the composite and generalization tests is markedly the least, as was found in the previous case.

It seems safe to say of the partial coefficients of the second order that there is highest correlation between the completion group and composite, and least between the composite and generalization group. It is again apparent that there is no evidence of identity of test results.

It will be seen later (page 53) that the composite here described correlates with the results of the Terman Group Test in 184 cases, .6812. When the results of the test-groups described in this section are correlated with the results of the Terman Group Test, the coefficients with a minor exception follow the same rank order.

CHAPTER VI

THE CONTROL GROUP

EFFECT OF FLUID TIME ALLOWANCE

In the administration of the tests of Selective and Relational Thinking and Generalization and Organization, the procedure which the writer has called *fluid time allowance* was followed. No attempt was made to require subjects to follow a prescribed time allowance for each test. Instead, their directions were: "Do Test 1 first; then do Test 2, then do Test 3, and so on, until you have finished. . . . If there is anything in any test that you cannot do, leave it. Go back to it later if you have time. Work as fast as you can, but do not make mistakes. . . . At certain times I shall say, 'Even if you have not finished Test 1 (or 2 or 3 or 4) begin Test 2 (or 3, or 4, or 5), etc.' This is to keep you from working too long on any one test. Do not wait for me to say it. As soon as you finish a test, go right ahead. . . ." ¹

It seemed worth while to attempt to determine whether this method of administration appreciably affected the correlations of test results as found in the preceding chapter. To study this problem, papers were especially prepared, as described on page 15, and directions for administering the tests (Appendix, pages 68-72) were so modified that subjects spent the time originally prescribed upon the single tests for which time was allowed.

The standard deviations and means of this examination for the one hundred eighty-four (184) cases to whom both Series A and B were administered are shown in Tables 18 and 19.

The intercorrelations of each of the content-groups and each of the form-groups were computed. These, together with the corresponding results secured from testing the miscellaneous

¹ See footnote, page 5.

TABLE 18
COMPARATIVE STANDARD DEVIATIONS AND MEANS OF GROUPS OF TESTS
DIFFERING IN CONTENT

Time-Control Group (184 Cases)

Groups of Tests	Series A (Jan. 10, 1924)		Series B (Mar. 7, 1924)	
	S.D.	Mean	S.D.	Mean
Words.....	24.0192	108.7821	24.6708	111.7170
Numbers.....	16.7888	61.8256	19.3008	67.1736
Space.....	8.5820	22.0216	10.8048	32.8040

TABLE 19
COMPARATIVE STANDARD DEVIATIONS AND MEANS OF GROUPS OF TESTS
DIFFERING IN FORM

Time-Control Group (184 Cases)

Groups of Tests	Series A (Jan. 10, 1924)		Series B (Mar. 7, 1924)	
	S.D.	Mean	S.D.	Mean
Completion.....	10.2364	40.3260	12.6390	45.1900
Analogies.....	11.0648	27.4128	11.9304	36.8912
Generalization.....	8.3586	22.1410	9.5178	24.4401
Composite.....	19.6960	95.3260	22.0552	96.4673

group of one thousand thirty-nine cases under the *fluid time allowance* procedure, are given in Table 20.

The probable error of the difference has been determined, and the quotient of the actual difference and the probable error of the difference is given in column six. There is variation between the two sets of coefficients, but variation must be expected because of the errors in sampling. It is apparent that most of the differences are positive, that is, the coefficients for the control group are slightly higher than those for the miscellaneous group. The tendency is explained in part, at least, by the larger standard deviations¹ within the control group; their existence would tend to increase the values of the correlation coefficients. It is probable, then, that had the two groups been

¹ Cf. Tables 5 and 13 with Tables 18 and 19.

TABLE 20

INTERCORRELATIONS OF TIME-CONTROLLED TEST (184 CASES) AND
COMPARISON WITH MISCELLANEOUS GROUP (1039 CASES)

"A" Indicates Series A 1922; "B" Indicates Series B 1923, and "Kom" Indicates a Composite

Group	Control (184 Cases)		Miscellaneous (1039 Cases)		P.E. (r_1-r_2)	(r_1-r_2) divided by P.E. (r_1-r_2)
	r_1	P.E. r_1	r_2	P.E. r_2		
Anal. A-Anal. B.	.7900	.0181	.7193	.0101	.0209	3.41
Anal. A-Kom. A.	.6592	.0279	.6164	.0127	.0306	1.39
Anal. A-Kom. B.	.6631	.0276	.5999	.0134	.0306	2.06
Anal. A-Com. A.	.5510	.0343	.4729	.0162	.0379	2.06
Anal. A-Com. B.	.5300	.0355	.5252	.0151	.0385	.12
Anal. A-Gen. A.	.5253	.0358	.4938	.0158	.0391	.75
Anal. A-Gen. B.	.5226	.0358	.4853	.0160	.0392	.95
Anal. A-Kom. A.	.6537	.0282	.6270	.0127	.0309	.86
Anal. B-Com. A.	.5841	.0325	.4999	.0157	.0360	2.33
Anal. B-Gen. A.	.5530	.0342	.4471	.0167	.0380	2.78
Gen. A-Gen. B.	.6110	.0309	.5643	.0143	.0340	1.37
Gen. A-Kom. A.	.4425	.0397	.4552	.0165	.0429	-.30
Gen. A-Kom. B.	.4248	.0404	.4214	.0172	.0439	.07
Gen. A-Com. A.	.4175	.0407	.3090	.0189	.0448	2.42
Gen. A-Com. B.	.3834	.0421	.3616	.0181	.0458	.47
Gen. B-Kom. A.	.4941	.0373	.4551	.0166	.0408	.95
Com. A-Kom. A.	.5817	.0327	.5449	.0147	.0358	1.02
Com. A-Kom. B.	.5834	.0325	.5207	.0153	.0359	1.74
Com. A-Com. B.	.6228	.0310	.5440	.0147	.0343	2.29
Com. B-Kom. A.	.5902	.0321	.5596	.0144	.0351	.87
Kom. A-Kom. B.	.7802	.0194	.7132	.0103	.0219	3.05
Words A-Words B.	.8301	.0153	.7660	.0086	.0175	3.68
Words A-Numbers A.	.5893	.0322	.5429	.0148	.0354	1.31
Words A-Numbers B.	.5631	.0337	.5703	.0149	.0368	-.48
Words A-Space A.	.5694	.0333	.4641	.0164	.0371	2.83
Words A-Space B.	.5347	.0352	.5048	.0156	.0385	.77
Words B-Numbers A.	.6175	.0305	.5516	.0146	.0348	1.89
Words B-Space A.	.5397	.0350	.4263	.0171	.0389	1.62
Numbers A-Numbers B.	.7281	.0286	.7084	.0104	.0304	.64
Numbers A-Space A.	.5670	.0335	.4599	.0165	.0373	2.87
Numbers A-Space B.	.5829	.0326	.5339	.0150	.0358	1.36
Numbers B-Space A.	.5791	.0328	.4617	.0164	.0366	3.20
Space A-Space B.	.6792	.0266	.6081	.0132	.0296	2.67
						53.02
Mean Difference.					1.606 P. E.	

TABLE 20 (Continued)

COEFFICIENTS OF CORRELATION OF TESTS GIVEN UNDER CONTROLLED
TIME (184 CASES)*Corrected for Attenuation*

Numbers—Space.....	.8261
Words—Numbers.....	.7585
Words—Space.....	.8262
Numbers—Space (Words).....	.6226
Words—Numbers (Space).....	.4254
Words—Space (Numbers).....	.2418
Completion—Composite.....	.8419
Analogies—Composite.....	.8356
Analogies—Completion.....	.8031
Analogies—Generalization.....	.7708
Generalization—Composite.....	.6636
Generalization—Completion.....	.6543
Completion—Composite (Generalization).....	.7205
Analogies—Composite (Generalization).....	.6789
Analogies—Completion (Generalization).....	.6188
Analogies—Generalization (Completion).....	.5444
Completion—Composite (Analogies).....	.5219
Analogies—Composite (Completion).....	.4961
Generalization—Composite (Completion).....	.2761
Generalization—Completion (Analogies).....	.0930
Generalization—Composite (Analogies).....	.0356
Composite—Completion (Generalization—Analogies).....	.5208
Composite—Analogies (Generalization—Completion).....	.4289
Composite—Generalization (Analogies—Completion).....	.0083

of equal variability, the differences between them would have been less. In only two cases, i.e., between generalization Series A and composite Series A, and word Series A and numbers Series B, the differences are negative; they are so small that they are regarded as chance variations only since both are less than .5 P.E. of the difference. In no case does the difference exceed what might reasonably occur by chance. When the entire series is considered, the mean difference is found to be 1.606 P.E. of the mean difference. It is possible, therefore, that the differences are those which may be regarded as strictly chance differences. However, the chances are greater that the difference in administration had some, but not a great deal, of effect upon the results of the test in such a way that the correlation coefficients have been affected.

EFFECT OF AGE

As described on page 14, the effect of age, as a factor affecting the correlations between tests of different types of content and with differences in form, has been controlled. In stating the results of computations based on the group of one thousand thirty-nine cases, as given on the preceding pages, there has been given also the corresponding value as computed for the group of four hundred eighty-nine boys of sixteen years and more but less than seventeen. The group, therefore, contains subjects whose chronological ages are limited to a range of one year. Intuitive judgment of the tables already examined leads one to conclude that there is little difference between the correlation coefficients computed for this group and those for the miscellaneous group. A more careful, but not thorough, analysis justifies the conclusion already suggested. In the cases below, arbitrarily chosen because of their importance, the differences between the coefficients for the two groups, expressed in the terms of the probable error of their differences, are:

	P.E.		P.E.
Words—Numbers.....	1.0020	Analogies—Completion.....	1.1028
Numbers—Space.....	.8989	Analogies—Generalization....	.2495
Words—Space.....	1.4036	Completion—Generalization..	.0669
Average Difference.....			.7872

Inasmuch as these representative cases disclose a mean difference of less than a unit probable error of the difference, the chances are more than equal that there is no real difference between them.

CONSTANCY OF CORRELATION COEFFICIENTS

The presence of errors in psychological measurements was recognized on page 27; Table 2 shows the reliability, or in popular language, the unreliability, of the tests which have been used. In the light of these two facts, the variability of correlations between pairs of traits should be examined briefly. The raw coefficients of correlation between the traits, studied in the second section of this chapter, will be used again as a representative and significant sampling. In each case, the four so-called cross correlations will be considered, i.e., the relation

between trait one of Series A and trait two of Series A and Series B, and between trait one of Series B and trait two of Series A and Series B. Table 21 gives the mean for each group of the absolute values of the deviations from their mean and mean of the probable errors of the four cross-correlation coefficients (Table 6).

TABLE 21

MEAN DEVIATIONS AND MEAN PROBABLE ERRORS OF CROSS-CORRELATION COEFFICIENTS

Groups of Tests	M.D.	M. of P.E.'s
Words—Numbers.....	.0132	.0144
Words—Space.....	.0127	.0167
Numbers—Space.....	.0473	.0155
Analogies—Completion.....	.0194	.0156
Analogies—Generalization.....	.0156	.0161
Completion—Generalization.....	.0165	.0186
Mean of Mean Deviations.....		.0208
Mean of Probable Errors.....		.0162

The technique of Table 21 is not entirely justifiable. However, since the differences are so small, the use of a mean of probable errors does not materially falsify the real facts, and it is a convenient method of roughly estimating the deviations. Four of the deviations fall within a single probable error of the mean value for the group; only one materially departs from it. The mean of the deviations from their mean is less than one and three-tenths of the mean of their probable errors. The deviations are therefore not regarded as significant.

The constant values of the cross-coefficients of correlation disclose another interesting fact. The reliability coefficients obtained between the results of Series A, 1922, and Series B, 1923, for the test grouped on the basis of content are (page 20): for words, .7660; for numbers, .7084; and for space, .6081. Clearly, there is evidence in these reliability coefficients of considerable shifts or changes in the relative standing of individuals in each of the three groups of tests. To determine the correspondence in these shifts in one set of results as compared with another, correlations, in each case, of the gains or differences

between the scores of Series A (1922) and Series B (1923) were computed. The results are:

	<i>r</i>	P.E.
Gain in Words with Gain in Numbers.....	.0218	.0209
Gain in Words with Gain in Space.....	.0350	.0208
Gain in Numbers with Gain in Space.....	.0881	.0208

It is evident that correlations between the gains in the respective content-groups approach zero, and this fact may be important in accounting for the constancy of value of the four cross-correlations in each case. In passing, one should notice that the existence of zero correlation between the gains is a necessary, but not an adequate, condition which must be met before the formula for correction for attenuation can be properly applied (page 29; also Yule, '22, page 214).

COMPARISON WITH TERMAN GROUP TEST RESULTS

Because of the value and general use of standardized tests of general intelligence, such as the Terman Group Test, the National Intelligence Tests, the Otis, the Miller, the Thorndike tests, and others, a definite comparison of the results of content-groups and form-groups with such a test was arranged. There are two particularly interesting points upon which information is desired:

First, is there gained by the result of a general intelligence test, equally good measures of one's ability to do the work of each of the content and form types considered, i.e., words, numbers, and spatial content; and analogies, completion, and generalization form?

Second, does the composite,¹ artificially constructed, as it was, for a rough evaluation of the completion, analogies, and generalization groups, perform its function of ranking them in respect to their correspondence with a standard and generally recognized test of general intelligence?

It was found that the Terman Group Test had been given throughout the grades in which the control tests were given, and its results, as it had been administered and scored by the

¹ Described on page 43.

school authorities, were accepted by the writer. The intercorrelations of these results with those of the I. E. R. test scores are shown in Table 22.

TABLE 22
CORRELATIONS BETWEEN SCORES ON TERMAN GROUP TEST AND I. E. R.
TEST SCORES

Groups of Tests	<i>r</i>	P.E.
Terman—Words.....	.7284	.0233
Terman—Numbers.....	.6042	.0316
Terman—Space.....	.5910	.0324
Terman—Words (Numbers).....	.5784	
Terman—Words (Space).....	.5910	
Terman—Words (Numbers—Space).....	.5419	
Terman—Numbers (Space).....	.4051	
Terman—Numbers (Words).....	.3161	
Terman—Numbers (Words—Space).....	.1628	
Terman—Space (Numbers).....	.3786	
Terman—Space (Words).....	.3132	
Terman—Space (Numbers—Words).....	.1568	

CORRELATIONS BETWEEN SCORES ON TERMAN GROUP TEST AND I. E. R.
TESTS DIFFERING IN FORM

Groups of Tests	<i>r</i>	P.E.
Terman—Composite.....	.6812	.0267
Terman—Analogies.....	.6788	.0268
Terman—Completion.....	.6274	.0302
Terman—Generalization.....	.4432	.0400
Terman—Analogies (Generalization).....	.5850	
Terman—Analogies (Completion).....	.5126	
Terman—Analogies (Generalization—Completion).....	.4639	
Terman—Completion (Generalization).....	.5431	
Terman—Completion (Analogies).....	.2534	
Terman—Completion (Generalization—Analogies).....	.2340	
Terman—Generalization (Completion).....	.2561	
Terman—Generalization (Analogies).....	.1406	
Terman—Generalization (Completion—Analogies).....	.0979	

It seems clear that the word-group correlates most highly with the Terman Group Test. The raw coefficient (for correlation for attenuation is impossible with the data at hand) of

the zero order is .7284 (P.I. .3148), and such coefficients have been regarded as high (Rugg, '17). When the Terman Group Test is correlated with numbers and space, the coefficients are .6042 (P.I. .3032) and .5910 (P.I. .1933), respectively.

When partial correlations were computed between each and the Terman test, the other two variables having been rendered constant, words correlates with the criterion .5419 (P.I. .1596); numbers, .1628 (P.I. .0233); and space, .1568 (P.I. .0123). In this, as in previous comparisons, the residuals show that none of the groups is identical with any other. It is equally clear that there is an overlapping of one upon the other. Finally, one can safely either say that word tests correspond highest with this measure of general intelligence, or that they have been weighted more heavily than other types of content in predicting efficiency of children in school work, for which the Terman test was especially constructed.

When the correspondence between the analogies, completion, and generalization tests and the Terman group test is studied, a somewhat similar range of variation is noticed. The coefficients of the zero order indicate a slightly higher correspondence between the Terman and the analogies group than between the former and the completion group. In both cases, the correspondence is greater than between the Terman and the group of generalization tests. With the other two rendered constant in turn, the correlation between Terman and analogies is .4639 (P.I. .1141); with completion, .2340 (P.I. .0278); and with generalization, .0979 (P.I. .0048).

The divergence between each of these from perfect correspondence is striking. There is, however, positive correspondence between each and the Terman group test; but it is not high, and in the case of generalization, the correspondence with the criterion is so slight that it seems to approach zero.

CHAPTER VII

CONCLUSION AND SUMMARY

The purpose of this chapter should be to interpret and summarize the results of the study. But since one tends to emerge more or less confused from the mass of detail presented, he is prompted at once to ask the question, "What does it mean?" The writer does not hope to answer the question in full. It may be that the interrelations which are significant are all more or less self-evident; however, it is entirely possible that there are, between the tests and groups of tests presented, important but subtle relations which are neither recognized nor valued. It is also true that a discussion of test results is more or less futile unless it is tacitly assumed, as in practical administration of all intelligence tests, that the trait or traits tested are sufficiently constant to warrant an attempt at measurement and that the tests used are at least sufficiently valid, adequate, and reliable measures to warrant their use. Upon these assumptions some rather definite conclusions seem to be warranted.

1. One is confronted first by the positive character of practically all of the correlations. The trend is one that has been noted and emphasized emphatically (Thorndike, '14 and others). There is absence of negative correlations in the interrelations of these more general traits (see page 4) as contrasted with the frequent occurrence of negative correlation between the narrower traits studied by McCall ('16).

2. A second significant fact is the range within which the values of the correlation coefficients fall. For the miscellaneous group of one thousand thirty-nine (1039) cases, the corrected coefficients, theoretically the truest (and highest) obtainable measures of correspondence one can secure, with available technique, range in value from .8666 to .4403; their predictive indices range from about .51 to .10. If pairs of scores were duplicates, their correlation should numerically approximate 1.00;

their deviations from this value imply differences between the results of the respective tests. While all combinations available have not been studied, an attempt has been made to include all typical correlations. All intercorrelations of the groups of tests have been computed, and samplings of intercorrelations of separate tests have been made. The latter, without exception, tend to correlate lower than the groups of tests, and further labor in that direction seems unjustified. In all the results no evidence is disclosed that the scores of the content-group of verbal, numerical and so-called spatial relations groups, or the groups of completion, analogies and generalization tests are duplicate and independent measures of a single function. Although the same general technique has been followed as used by him, these results are opposed to the earlier views¹ of Spearman ('04 A, page 273) who, speaking of universal unity of intellectual functions, says:

In view of this community being discovered between such diverse functions as in-school Cleverness, outdoor Common Sense, Sensory Discrimination and Musical Talent, we need scarce be astonished to continually come upon it no less paramount in other forms of intellectual activity. Always in the present experiments approximately

$$\frac{r_{pq}}{\sqrt{r_{pp} \times r_{qq}}} = 1.00.$$

I have actually tested this relation in twelve pairs of such groups taken at random, and have found the average value to be precisely 1.00 for the first two decimal places with a mean deviation of only 0.05. All examinations, therefore, in the different sensory, school, or other specific intellectual faculties, may be regarded as so many independently obtained estimates of the one great common Intellectual Function.

Though the range of this central function appears so universal and that of the specific functions so vanishingly minute, the latter must not be supposed to be altogether non-existing. We can always come upon them eventually, if we sufficiently narrow our field and consider branches of activity closely enough resembling one another.

That narrowness is not characteristic of the traits measured by the tests of Selective and Relational Thinking and Generalization and Organization is evidenced by the variety of tests which they include: arithmetical problems; absurdities; line arrangements; opposites; number series completion; spatial,

¹ A discussion of the relation of these findings to Professor Spearman's more recent views is not attempted here.

grammatical, and verbal analogies; moral judgments; verbal, numerical, and spatial generalization; sentence completion; cutting up surfaces; and disarranged sentences. Such a battery does not suggest the testing of a narrow function.

It should be borne in mind that this compilation of tests was made not for the present study, but as a useful measure of general improvement in mental abilities (page 4). Thorndike ('22) describes the tests as measures of "ability to think with symbols, to understand and apply generalization, to discern and use relations, and to select essential facts or elements, and to organize facts for a purpose — in James' phrase to 'think things together.'" It appears, therefore, that these tests are not tests of a narrow field, and that the results do oppose the early findings of Spearman as quoted.

The range of the partial coefficients of the first order is also significant. No coefficient of the first order is as low as zero. The nearest approach to this is the relation of the generalization tests to the completion tests when analogies are rendered constant, in which case it is $.0573 \pm .0209$. The next nearest approach to zero found among the correlations of test-groups is $.1823 \pm .0203$, between the generalization tests and the composite, when analogies are rendered constant. This and all other correlation coefficients between test-groups are safely above any probability of being chance deviations from zero. The same tendencies are shown in a study of the raw scores and also in the intercorrelations of the single tests.

This fact is not inconsistent with the implications of the failure of corrected coefficients to approach unity. For, if either of the variables between which correlation is computed is identical with one which has been rendered constant, the resulting correlation is zero (Yule, '22, page 235). This result has not been obtained in any case, and while certain conditions make it undesirable to attempt to draw definite conclusions from the fact that coefficients of the first order are positive and do not equal or approximate zero, the results are at least not inconsistent with those previously found.

A third result suggests the same conclusion. It was found (page 51) that the four cross-correlations between the successive pairs of scores, without exception, are practically equal. The mean deviation, from their mean, of the cross-correlations

for the content- and form-groups is .0208, and the mean probable error is .0162 (page 51). From the same data it was found that the raw intercorrelations between the gains in the various content-groups as shown by scores of Series B (1923) when compared with Series A (1922) approximate zero. The correlation between gain in words and gain in numbers is $.0218 \pm .0209$; between gain in words and gain in space — $.0350 \pm .0208$; and between gain in numbers and gain in space, $.0881 \pm .208$. If the measures are duplicate measures of a common "intellective function," the gain in one should parallel the gain in another and the correlation of gains between the two should approach unity. It is recognized that these are uncorrected coefficients and that their reliabilities, as computed, are low; if corrected coefficients could be obtained, the correlations might be found to be substantial. However, as they stand, the raw coefficients approximate zero, and independence between gains is suggested, rather than that the tests are duplicate measures.

It has been found, first, that without exception in the large number of correlations computed no coefficient of the zero order reaches unity; second, that no coefficient of the first order reaches zero; and third, that correlation between gains in these respective scores is very low. Each of these three independently derived results suggests the same implication that the three content-groups of verbal, numerical, and so-called spatial relations are not duplicate measures of the same function or functions; and, similarly, that the completion, analogies, and generalization tests are not duplicate measures. From a different point of view, the same conclusion might be stated: Differences in the relative standings of pupils occur when they are given so-called intelligence examinations having differences either in form or in content. It appears that the inference, if supported by additional studies, is tremendously significant for makers and users of tests.

3. Although no evidence has been found to indicate that the test results here studied are duplicate and independent measures of the same trait or function, their persisting positive concomitant variation has been shown. The amount of correlation varies with the specific tests considered, for a statement of which the reader is referred to the preceding chapters. Variability of coefficients has been found commonly in correlation studies, but

is perhaps best seen in Yerkes' ('21, pages 573-680) voluminous description of the extensive psychological examining done in the United States Army. The variations in values of the correlation coefficients of each of the zero, first, and second orders found here imply that the interrelations of the tests are very complex.

4. Pages 46 to 49 described the results of the *fluid time allowance* method of test administration. To that discussion little needs to be added. It is evident that it is possible to arrange tests mechanically so that subjects will be little disposed to work on tests other than the one for which time is allotted; but in the present study, where the material seems ample to challenge and fill the time of the subjects, the difference in method of administration had little, and possibly no, effect upon the intercorrelations of the respective test results.

5. The effect of age, also, within the limits of the present study, has been described (page 50). The intercorrelations of scores of boys, all of sixteen years or more but less than seventeen, are approximately equal to those of the miscellaneous group. There is, therefore, at this age little, if any, correlation between chronological age and the results of the tests. The mean difference between the coefficients is $\pm .7872$ P.E. of their difference and therefore can be regarded as due to errors of sampling.

6. There is interesting variation in the correlation of the various content- and form-groups with such a recognized general intelligence test as the Terman Group Test. The word-group correlates with the Terman Group Test .7284 (P.I. .3148); the number-group correlates with the same criterion .6042 (P.I. .2032); and the space-group, .5910 (P.I. .1933). These, however, are uncorrected coefficients, for only one measure by the Terman Group Test of these subjects has been obtained. The figures, therefore, are probably too low (see page 29) to represent theoretical values. If any implications, however, can be drawn from the raw coefficients and if the results of the study have any general significance, it is evident that so-called general intelligence tests are dominantly tests which correspond to tests requiring verbal ability; and that there is considerably less association between the scores in so-called general intelligence tests and those of tests of number and space content.

Other studies have disclosed the same tendency. Of the general intelligence tests, it is probable that the Binet-Simon test in its various modifications is best known. Of the original test Burt ('11, page 180) says: "*Linguistic ability and linguistic attainments exert upon the Binet-Simon tests a special and positive influence of their own.*" And again, "Numerous factors affect the measurement of a child's intelligence by means of the Binet-Simon Scale . . . educational, and particularly linguistic attainments more profoundly than any other factor measurable with exactitude."

Terman ('16, page 230), speaking of the Stanford Revision of the Binet-Simon tests, says: "Our statistics show that in a large majority of cases the vocabulary test alone will give us an intelligence quotient within 10 per cent of that secured by the entire scale."

7. Variation is found also in the amount of correspondence between the form-groups and the Terman Group Test. With the latter, the composite (see page 53) correlates most highly, as might have been guessed, but the analogies and completion tests follow almost within a single possible error of sampling in the order given. Wyatt ('13) found the same: "The analogies and completion tests give the highest correlations with the subjective estimations of intelligence. The intercorrelations of each of these with every other are also high."

The generalization tests correlate with the Terman Group Test only .4432 (P.I. .1036), which indicates that it does not measure well those traits which heretofore have been considered as general intelligence. Or, stating the same fact reversely, the Terman Group Test does not measure equally well the abilities to do work in verbal, numerical, and spatial tests, or analogies, completion and generalization tests.

From a study of *Measurements of Mechanical Ability*, the content of the tests used dealing with mechanical content, Stenquist ('23, page 85) found that "An individual's position in General Intelligence is thus shown to be largely independent of his position in General Mechanical Ability and Aptitude," and (page 82) "The tests of mechanical ability herein described serve as an example and case in point, showing a type of intelligence and also emphasizing the need for clearer definition of just what we mean when we say a child has but little intelligence."

If studied from the viewpoint of predictive value of the uncorrected coefficients, the correlation between the respective content- and form-groups with the Terman Group Test would reduce, if so used, the error of optimum chance prediction (page 31), only from about ten per cent to a maximum of less than thirty-two per cent. Burt ('21, page 206) also found that "with the exception of absurdities and mixed sentences the value of any single test as a criterion of normal intelligence proves singularly low." The value of general intelligence tests has been demonstrated, but the limitations of their use for prediction of scores in certain tests need no further comment here.

The problem of test-making has not yet been solved; the value of special aptitude tests and tests of special abilities is evident; the utility of a more general measure has proved also to be of extreme importance. How accurately measures of abilities in different traits can be predicted from a single general measure is a problem for specialists in that field. Differences in form and content of tests are two elements which must be always considered, but the value of the respective types of each for special purposes remains to be demonstrated empirically.

Of this whole problem, Burt ('21, page 74) well says: "In abstracting from . . . scales a single measure for mental efficiency as a whole—for general intelligence, as it is commonly termed—each test must be weighted, not equally or arbitrarily, but according to an empirical 'regression coefficient' based upon its special correlation with intelligence itself. . . . Only by such a scale, or such a system of scales, can we diagnose general ability with scientific exactitude. But a scheme so elaborate will demand for its completion many years of coöperative research."

8. Finally, the predictive index is suggested as a measure of the significance of a correlation coefficient from the viewpoint of prediction. It numerically states the significance of a correlation coefficient in terms of the variability of the trait itself. For purposes of prediction it gives a definite, meaningful, and unambiguous measure of the value of an r . The predictive index is the arithmetical complement to the coefficient of alienation (Kelley, '23, page 216). While the latter gives the value of r by stating how much variability in prediction is not accounted for by the correlation coefficient, the predictive index

gives a positive statement of the ratio by which error has been reduced by means of prediction from the regression equation.

When correlation coefficients are so evaluated, the need for further study and analysis of tests whereby higher correlations, including reliability coefficients, might be obtained is obvious. The practical value for predictive purposes of coefficients now commonly obtained in psychological and educational measurements is too small to be considered a satisfactory tool for most effective use in prognosis and advisement.

SUMMARY

The following summary of conclusions is made:

1. The correlations throughout are positive. This holds between groups of tests when they are combined on the basis of content, verbal, numerical, or spatial; or on the basis of form, completion, analogies, or generalization. It is true of the relation between the various types of tests and the Terman Group Test. There is also consistently positive, but smaller, correlation between the results of the single tests studied.

2. No adequate evidence has been found to indicate that the results of the tests differing in type of content, such as words, number, and space, and differing in form, such as analogies, completion, and generalization, are duplicate and independent measures of a common mental function. These results are opposed to the early findings of Spearman ('04 A). On the other hand, it has been found that differences in the relative standings of pupils occur when they are given tests having differences either in form or in content.

3. In this study, the original use of the fluid time allowance as contrasted with the fixed time method (see page 46) had very little, and possibly no, effect upon the intercorrelations of the relative test results.

4. When the effect of chronological age was rendered constant at sixteen years, the mean difference of intercorrelations of the content-groups and the form-groups between the age-control group and the miscellaneous group of one thousand thirty-nine (1039) cases, is .7872 P.E. of their difference. It appears, therefore, that the effect of change of age, at this level, has little, if any effect upon relative test results.

5. Although the duplicate forms of the tests were administered at an interval of about one year, there is marked constancy in the intercorrelations of the test results. While the mean of their probable errors is .0162, the mean deviation of the absolute values of the coefficients of intercorrelation of the results for the content- and form-groups is .0208 (page 51).

6. The raw correlations between gains made in Series B, given in 1923, over the scores made in Series A, given in 1922, as determined for the content-group, are very low and approximate zero (page 52).

7. The correlations (1039 cases) between the respective content-groups and the Terman Group Test range from .5910 to .7284; between the form-groups and the same criterion, the correlations range from .4432 to .6812 (Table 21).

8. The predictive index is suggested as a measure of the significance, from the predictive viewpoint, of a correlation coefficient. Its meaning is more definite and its significance less ambiguous and less susceptible to misinterpretation than an r . The need for further study and analysis of tests, whereby higher correlations, including reliability coefficients, might be obtained is obvious in the restricted value in prediction of correlation coefficients now commonly obtained in psychological and educational measurements.

APPENDIX

I

TABLE 23

TRANSMUTATION TABLE I*

GROUPING TABLE FOR DETERMINING TRANSMITED SCORES FROM RAW
SCORES ON GROUPS OF TESTS

Group	Word Group	Number Group	Space Group	Completion Group	Analogies Group
0	†18-26.99	8.15.99	0-3.99	0-4.99	0-3.99
1	27-35	16-23	4-7	5-9	4-7
2	36-44	24-31	8-11	10-14	8-11
3	45-53	32-39	12-15	15-19	12-15
4	54-62	40-47	16-19	20-24	16-19
5	63-71	48-55	20-23	25-29	20-23
6	72-80	56-63	24-27	30-34	24-27
7	81-89	64-71	28-31	35-39	28-31
8	90-98	72-79	32-35	40-44	32-35
9	99-107	80-87	36-39	45-49	36-39
10	108-116	88-95	40-43	50-54	40-43
11	117-125	96-103	44-47	55-59	44-47
12	126-134	104-111	48-51	60-64	48-51
13	135-143	112-119	52-55	65-69	52-55
14	144-152	120-127	56-59	70-74	56-59
15	153-161	128-135	60-63	75-79	60-63
16	162-170	136-143	64-67	80-84	64-67
17	171-179	144-151	68-71	85-89	68-71

Step	Generalization Group	Composite Group	Totals (All Tests)	Terman Group
0	0-2.99	20-29.99	50-65.99	20-29.9
1	3-5	30-39	66-81	30-39
2	6-8	40-49	82-97	40-49
3	9-11	50-59	98-113	50-59
4	12-14	60-69	114-129	60-69
5	15-17	70-79	130-145	70-79
6	18-20	80-89	146-161	80-89
7	21-23	90-99	162-177	90-99
8	24-26	100-109	178-193	100-109
9	27-29	110-119	194-209	110-119
10	30-32	120-129	210-225	120-129
11	33-35	130-139	226-241	130-139
12	36-38	140-149	242-257	140-149
13	39-41	150-159	258-273	150-159
14	42-44	160-169	274-289	160-169
15	45-47	170-179	290-305	170-179
16	48-50	180-189	306-321	180-189
17	51-53	190-199	322-337	190-199

* The data from which these cases have been selected and from which computations have been made are on file in the Division of Psychology, Institute of Educational Research, Teachers College.

† The figures in the left-hand column of each pair indicate the lower limit of the steps.

TABLE 24
TRANSMUTATION TABLE II
GROUPING TABLE FOR DETERMINING TRANSMUTED SCORES FROM RAW
SCORES OF COMPONENT TESTS

Step	Test I	Test II	Test III	Test IV	Tests V to X
0	*0-3.99	-8--6.01	0-.99	0-2.99	0-1.99
1	4-7	-6--4.01	1	3-5	2-3
2	8-11	-4--2.01	2	6-8	4-5
3	12-15	-2--.01	3	9-11	6-7
4	16-19	0-1.99	4	12-14	8-9
5	20-23	2-3	5	15-17	10-11
6	24-27	4-5	6	18-20	12-13
7	28-31	6-7	7	21-23	14-15
8	32-35	8-9	8	24-26	16-17
9	36-39	10-11	9	27-29	18-19
10	40-43	12-13	10	30-32	20-21
11	44-47	14-15	11	33-35	22-23
12	48-51	16-17	12	36-38	24-25
13	52-55	18-19	13	39-41	26-27
14	56-59	20-21	14	42-44	28-29
15	60-63	22-23	15	45-47	30-31
16	64-67	24-25	16	48-50	32-33
17	68-71	26-27	17	51-53	34-35

Step	Tests XI and XII	Test XIII	Test XIV	Test XV
0	0-.99	0-2.99	0-.99	0-2.99
1	1	3-5	1	3-5
2	2	6-8	2	6-8
3	3	9-11	3	9-11
4	4	12-14	4	12-14
5	5	15-17	5	15-17
6	6	18-20	6	18-20
7	7	21-23	7	21-23
8	8	24-26	8	24-26
9	9	27-29	9	27-29
10	10	30-32	10	30-32
11	11	33-35	11	33-35
12	12	36-38	12	36-38
13	13	39-41	13	39-41
14	14	42-44	14	42-44
15	15	45-47	15	45-47
16	16	48-50	16	48-50
17	17	51-53	17	51-53

* The figures in the left-hand columns indicate the lower limit of the steps.

TABLE 25
 REPRESENTATIVE CORRELATION TABLE
 WORDS (SERIES A) AND NUMBERS (SERIES B)
 NUMBER SERIES B (1923)

Steps *	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	Total
17												1			1	1			3
16							1							1					2
15						1	1	2	1	2	1	3	1	3	2		1		18
14								1	2	6	8	6	8	3	2	2			38
13						1	4	8	5	13	4	6	4	3	2				50
12			1			3	7	11	16	19	22	10	2	3					94
11		1	1		1	4	10	17	23	27	15	15	6	3					123
10				1	7	7	12	23	25	27	14	6	3						125
9			1	3	6	8	23	25	22	22	8	5	3	1				1	128
8				3	9	15	27	34	30	16	12	7	2	1					156
7	1			2	12	14	26	27	14	10	4								110
6			2	7	6	15	23	13	7	6	4								83
5		2	2	5	5	15	16	8	3	3	2	1	1						63
4			2	4	3	5	3	6	2	4	1	1							31
3				1	2	2	2	1	1										9
2	1	1		1								1							4
1			1					1											2
0																			
Total	2	4	10	27	51	90	155	177	151	155	95	62	30	18	7	3	1	1	
																			1039

Coefficient of Correlation, Pearson Product-Moment † (Toops Formulæ)... .5703

Correlation Ratio of Words on Numbers †..... .5822

The difference is 2.71 times the probable error of the difference.

* For step values, see Transmutation Table I, page 65.

† Sheppard's Correlation not applied

II

*MODIFIED INSTRUCTIONS FOR TIME-CONTROLLED TEST

INSTRUCTIONS FOR GIVING THE I. E. R. TESTS OF SELECTIVE AND RELATIONAL THINKING AND OF GENERALIZATION AND ORGANIZATION

The material required for the examiner is a watch with a second hand.

The material required for each pupil is:

- 1 test booklet, Practice Form. This serves for both tests.
- 1 test booklet, Selective and Relational Thinking.
- 1 test booklet, Generalization and Organization.
- 2 pencils.

The time required is: Practice Form, 20 minutes; Selective and Relational Thinking, 39 minutes; Generalization and Organization, 36 minutes—95 minutes.

Twenty (20) minutes or less should suffice for oral directions, distribution of the tests, and collecting them. The total time will thus be less than two hours.

Have the pupils seated with desks clear of everything save two pencils and two or three blank sheets of paper. In case they do not have blank sheets of paper, tell them that they can figure on the margins of the pages and on the blank pages that will be found in the tests. Say:

"We are going to give you some tests to see how well you can think. The first will be a practice test and will not count on your score. The second and third will be the real tests. Do not ask any questions. Listen to the directions that I give, and do the best you can. We are going to give each of you a booklet like this (hold up a Practice Form) for a practice trial to show you what the examination is like.

"Do not begin work until I say, 'Go.' For the practice test, it is not necessary for you to use your pencils. You may do so if you wish. Since this is a practice test, it is important that you study the material to see just what you are to do.

"When I say, 'Go,' study Test A. Work each of the items carefully. If there is any item that you cannot do in a short time, pass on to the next and come back to it later if you have time. When I give you the signal we shall leave Test A and pass on to Test B, etc. Study each test as we come to it. Ask no questions. Do what you think is right.

"All ready to begin Test A.—Go!"

Time ninety (90) seconds.

"Stop!—Leave Test A. Ready to begin Test B.—Go!"

Time ninety (90) seconds.

"Stop!—Leave Test B. Ready to begin Test C.—Go!"

Time ninety (90) seconds.

"Stop!—Leave Test C. Ready to begin Test D.—Go!"

Time ninety (90) seconds.

"Stop!—Leave Test D. Ready to begin Test E.—Go!"

Time ninety (90) seconds.

"Stop!—Leave Test E. Ready to begin Test F.—Go!"

Time ninety (90) seconds.

* These instructions are adapted from the original instructions for giving the I. E. R. Tests of Selective and Relational Thinking and Generalization and Organization. See footnote, page 5.

"Stop!—Leave Test F. Ready to begin Test G.—Go!"

Time sixty (60) seconds.

"Stop!—Leave Test G. Ready to begin Test H.—Go!"

Time sixty (60) seconds.

"Stop!—Leave Test H. Ready to begin Test I.—Go!"

Time ninety (90) seconds.

"Stop!—Leave Test I. Ready to begin Test J.—Go!"

Time ninety (90) seconds.

"Stop!—Leave Test J. Ready to begin Test K.—Go!"

Time ninety (90) seconds.

"Stop!—Leave Test K. Ready to begin Test L.—Go!"

Time ninety (90) seconds.

"Stop!—Leave Test L. Ready to begin Test M.—Go!"

Time ninety (90) seconds.

"Stop!—Leave Test M. Ready to begin Test N.—Go!"

Time ninety (90) seconds.

"Stop! Pencils down!"

Have the books collected quickly.

After the Practice Form booklets have been collected, continue:

"Now we shall have the regular examination. I shall give each of you a booklet like this (holding up a booklet of *Selective and Relational Thinking*). When you receive it, read what it says on the front cover. Do not break the seals or open it until I tell you to. Read what it says on the front cover; write your name, and age, and grade, and the subjects you are studying this year. Then sit quietly until I give you instructions what to do next. When I tell you to, raise the upper corner of the first printed page, insert your pencil (illustrating), and carefully tear loose the first page from its fastening at the right. Then fold your books back so that Test 1 will be on top and the booklet will form a pad on which to write. Do not begin until I say, 'Go.' Work through the test carefully. If you finish it before I say, 'Stop,' go back over your work and make any changes or corrections that you may wish to make. Work as fast as you can, but do not make mistakes. Work steadily; do not get excited or worried.

"Remember, you are not to break any seal or turn any page until you are told to. Work on each test during the full time that is allowed for it. If you finish it, go back and perfect your work on that test. Never work before or back of the test on which the whole class is working. Do not ask any questions. Do what you think is right."

Distribute the *Selective and Relational Thinking Booklets*. When the booklets have been distributed, say:

"First, print your name. Print your first name first, then, your last name. Write your age in years and months. Write the number and section of your school grade. Write the names of the subjects you are studying this year."

When all have filled in the blanks on the first page, continue:

"Now raise the upper corner of the first printed page; insert your pencil just below this page, but above the blank cover-page beneath. Carefully tear the first page loose from its fastening at the right. Fold your books back so that

Test 1 will be on top and the booklet will form a pad on which to write. Do not begin until I say, 'Go.' When you are ready, hold your pencils up."

While they are adjusting their papers, continue:

"Remember, work as fast as you can, but do not make mistakes. If you finish a test before I say, 'Stop,' go back over that test and perfect your work, but never work on a test before or back of the one on which the whole class is working."

When all have found Test 1, say:

"Ready,—Go!" Allow *ten (10) minutes*, then promptly say:

"Stop! Pencils up!—Turn the booklet over. Raise the upper right-hand corner of the cover-page. Insert your pencil just beneath this page, but above the page which is numbered Test 2. Carefully tear the cover-page loose from its fastening at the right, and fold it back so that Test 2 will be on top and the booklet will form a pad on which to write. When you are ready hold your pencils up. Do not begin until I say, 'Go.' Remember, if you finish a test before I say, 'Stop,' go back over it and correct it, but do not turn to any other page. Spend your time on the test on which the whole class is working."

When all are ready for Test 2 say:

"Ready, Go!" Allow *four (4) minutes*, then promptly say:

"Stop! Pencils up! Now raise the upper right-hand corner of the page to Test 3. Insert your pencil between this page and the sheet covering Test 5. Be careful while loosening this page; keep the seal of Test 4 unbroken. Loosen the fastening of this page at the right, and fold your books back. Test 3 should be on top; Test 4 should be covered. When you are ready, hold your pencils up."

When all are ready, say:

"Ready, Go!" Allow *four (4) minutes*, then say promptly:

"Stop! Pencils up. Now insert your pencils beneath the cover-sheet of Test 4, just below where you have been working. Loosen the cover-sheet and fold it back or tear it out. Open your books, for a part of Test 4 is on the following page. Do you all see that a part of Test 4 is on the following page?"

When all are ready, say:

"Ready, Go!" Allow *five (5) minutes*. Then promptly say:

"Stop! Pencils up! Now insert your pencils beneath the cover-sheet of Test 5. Carefully loosen it from its fastening at the right, then fold it back so that Test 5 will be on top and the booklet forms a pad on which to write."

When all are ready, say:

"Ready, Go!" Allow *four (4) minutes*. Then promptly say:

"Stop! Pencils up! Now raise the upper right-hand corner of the printed page. Insert your pencil beneath it, but above the cover-sheet of Test 7. Carefully loosen the fastening, and fold the leaf over so that Test 6 will be on top, and the booklet will form a pad on which to write." When all are ready say:

"Ready, Go!" Allow *four (4) minutes*. Then promptly say:

"Stop! Pencils up! Turn your booklets over. Raise the upper right-hand corner of the cover-sheet; insert your pencil and carefully loosen the fastening. Fold back the cover-sheet so that Test 7 is on top and the booklet forms a pad on which to write."

When all are ready, say:

"Ready, Go!" Allow *four (4) minutes*. Then promptly say:

"Stop! Pencils up! Turn your booklets over. Insert your pencil beneath

the cover-page and loosen it from its fastening. Fold it back, so that Test 8 will be on top, and the booklet will form a pad on which to write."

When all are ready, say:

"Ready, Go!" Allow *four (4) minutes*. Then promptly say:

"Stop! Pencils up!" Have the booklets collected at once. Pieces of scratch paper should not be collected.

(Part II)

Then say:

"Next we shall give each of you a booklet like this (holding up the Generalization and Organization booklet). We will do this in the same way. When you receive it, write your name and age and grade, the section you are in and your studies; but do not open the booklet or break any seals until I tell you to."

Have the Generalization and Organization booklets distributed quickly. Allow about two minutes for writing name, age, grade, section and subjects. When all have filled in the blanks on the first page, say:

"Raise the upper corner of the first printed page; insert your pencil just beneath this page, but above the blank cover page beneath. Carefully tear the first page loose from its fastening at the right. Fold your books back so that Test 1 will be on top and the booklet will form a pad on which to write. Do not begin until I say, 'Go.' When you are ready hold your pencils up."

When all have found the page and are ready, say:

"Ready, Go!" Allow *two (2) minutes*. Then promptly say:

"Stop! Pencils up! Turn your booklets over. Insert your pencil beneath the upper right-hand corner of the cover-sheet and loosen it from its fastening. Fold it back so that Test 2 will be on top and the booklet will form a pad on which to write. When you are ready hold your pencils up."

When all are ready, say:

"Ready, Go!" Allow *three (3) minutes*. Then promptly say:

"Stop! Pencils up! Insert your pencil beneath the upper corner of the page. Carefully tear the page loose from its fastening and turn your books back so that Test 3 will be on top and the booklet will form a pad on which to write. When you are ready hold your pencils up."

When all are ready, say:

"Ready, Go!" Allow *three (3) minutes*. Then promptly say:

"Stop! Pencils up! Turn your booklets over. Insert your pencils beneath the upper corner of the cover-sheet and loosen it. Fold your papers in the usual way. Test 4 should be on top."

When all are ready, say:

"Ready, Go!" Allow *three (3) minutes*. Then promptly say:

"Stop! Pencils up! Insert your pencils just beneath the upper right-hand corner of this page, but above the cover-sheet beneath. Loosen the page and fold it back in the usual way. Test 5 should be on top." When you are ready, hold your pencils up."

When all are ready, say:

"Ready, Go!" Allow *ten (10) minutes*. Then promptly say:

"Stop! Pencils up! The directions for the next test are just below where you

have been working. The figures, however, are on the next page. Loosen the cover-sheet and tear it out or fold it back. When you are ready hold your pencils up." When all are ready, say:

"Ready, Go!" Allow *three (3) minutes*. Then promptly say:

"Stop! Pencils up! Insert your pencils beneath the upper right-hand corner of this page and loosen it from its fastening. Fold it back in the usual way. Test 7 should be on top. When you are ready hold your pencils up."

When all are ready, say:

"Ready, Go!" Allow *twelve (12) minutes*. Then promptly say:

"Stop! Pencils up!"

Have the booklets collected at once. Later, have the loose sheets collected and destroyed.

Fill in the blanks below and place this sheet on top of the test blanks. Tie the papers for the group together so that they will not be confused with other papers.

Name of School.....

Number of Grade.....

Section of Grade.....

Date when test was given.....

Notation of any unusual or important interruptions, distractions, failure to comply with instructions, etc.:

Name of Examiner.....

III

BIBLIOGRAPHY

- BALLARD, P. B. ('20): *Mental Tests*. Hodder and Stoughton, London, England.
- BURT, C. ('21): *Mental and Scholastic Tests*. London County Council, London, England.
- GARNETT, J. C. MAXWELL ('21): *Education and World Citizenship*. Cambridge University Press.
- GATES and LA SALLE ('23): "The Relative Predictive Values of Certain Intelligence and Educational Tests, Together with a Study of the Effect of Educational Achievement Upon Intelligence Scores." *Jour. of Educ. Psych.*, Vol. XIV, No. 9, pp. 517-539.
- HART, B. and SPEARMAN, C. ('12): "General Ability, Its Existence and Nature." *British Jour. of Psych.*, Vol. V, pp. 51-84.
- KELLEY, TRUMAN L. ('23): *Statistical Method*. The Macmillan Company.
- MCCALL, WM. A. ('23): *How to Experiment in Education*. The Macmillan Company.
- MCCALL, WM. A. ('16): *Correlation of Some Psychological and Educational Measurements*. Contributions to Education, No. 79, Teachers College, Columbia University.
- MINER, JOHN RICE ('22): *Tables of $\sqrt{1-r^2}$ and $1-r^2$ for Use in Partial Correlation and Trigonometry*. Johns Hopkins Press.
- PINTNER, RUDOLF ('23): *Intelligence Testing*. Henry Holt and Company.
- RUGG, HAROLD O. ('17): *Statistical Methods Applied to Education*. Houghton Mifflin and Company.
- SEASHORE, CARL E. ('19): *The Psychology of Musical Talent*. Silver, Burdett and Company.
- SPEARMAN, C. ('04 A): "'General Intelligence' Objectively Determined and Measured." *American Jour. of Psych.*, Vol. XV, pp. 200-293.
- SPEARMAN, C. ('04): "The Proof and Measurement of Association Between Two Things." *American Jour. of Psych.*, Vol. XV, pp. 72-101.
- SPEARMAN, C. ('23): *The Nature of Intelligence and the Principles of Cognition*. The Macmillan Company.
- STERN, W. ('14): *The Psychological Method of Testing Intelligence* (Tr. Whipple, G. M.). Educational Psychology Monographs, No. 13. Warwick and York, Baltimore.
- SYMPOSIUM ('21): "Intelligence and Its Measurement." *Jour. of Ed. Psych.*, Vol. XII, Nos. 3 and 4.
- THOMSON, GODFREY H.: "General versus Group Factors in Mental Activities." *Psych. Rev.*, Vol. XXVII, No. 3, pp. 173-190.
- THOMSON, GODFREY H. ('23): "On Hierarchical Order Among Correlation Coefficients." *Biometrika*, Vol. XV, pp. 150-160.
- THOMSON, GODFREY H. ('21): "On the Degree of Perfection of Hierarchical Order Among Correlation Coefficients." *Biometrika*, Vol. XII, pp. 355-366.
- THOMSON, GODFREY H. ('20): "The General Factor Fallacy in Psychology." *British Jour. of Psych.*, Vol. X, pp. 319-326.

- THORNDIKE, E. L. and WOODWORTH, R. S. ('08): "The Influence of Improvement in One Mental Function Upon the Efficiency of Other Functions." *Psych. Rev.*, Vol. VIII, pp. 217-26 (1901).
- THORNDIKE, E. L. ('14): *Educational Psychology*, Vol. III. Teachers College, Columbia Univ., N. Y.
- THORNDIKE, E. L. ('22): "Instruments for Measuring the Disciplinary Values of Studies." *Jour. of Ed. Research*, Vol. V, No. 4, pp. 269-279 (April, 1922).
- THORNDIKE, E. L. ('24): "Mental Discipline in High School Studies." *Jour. of Ed. Psych.*, Vol. XV, No. 1, Jan., 1924, pp. 1-22.
- THORNDIKE, E. L. ('24 B): "Mental Discipline in High School Studies." *Jour. of Ed. Psych.*, Vol. XV, No. 2, pp. 83-98 (Feb., 1924).
- TOOPS, HERBERT A. ('21): "Eliminating the Pitfalls in Solving Correlation," A Printed Correlation Form. *Jour. of Exp. Psych.*, Vol. IV, No. 6 (Dec., 1921).
- WISSLER, CLARK ('01): "The Correlation of Mental and Physical Tests." *Psych. Rev. Monograph, Supplement*, No. 16.
- WYATT, STANLEY ('13): "The Quantitative Investigation of Higher Mental Processes." *British Jour. of Psych.*, Vol. VI, pp. 109-133.
- YERKES, ROBERT M. (Editor) ('21): *Psychological Examining in the United States Army*. Memoirs of the National Academy of Science, Vol. XV. Government Printing Office, Washington, D. C.
- YULE, G. UDNY ('22): *An Introduction to the Theory of Statistics*. Charles Griffin and Company, Ltd., London, England.

VITA

Edwin Maurice Bailor was born in Nebraska, May 13, 1890. His elementary, high school, and early college training were received in Iowa. In 1914 he received the degree of Bachelor of Arts in Education from Washington State College, and in 1916 was granted the degree of Master of Arts in Education by the same institution.

From 1909 to 1913, and from 1914 to 1915, he was engaged in teaching, supervision, and administration of rural, village, and consolidated schools in Lewis County, Washington. From 1915 to 1918 he was Instructor in Education and Psychology at Washington State College; from 1918 to 1919, he served in the Sanitary Corps, U. S. Army as psychological examiner and supervisor of reconstruction work; from 1919 to 1921 he was civilian psychologist, in charge, directing vocational guidance for the Motor Transport Corps, Quartermaster Department, U. S. Army; and from 1921 to 1923 he was supervisor of commercial and correspondence training for District 4 of the U. S. Veterans' Bureau, located in Washington, D. C. From February, 1923, to June, 1924, he was Assistant in Educational Psychology, Teachers College, Columbia University.

